



Helsinki
Center
of
Economic
Research

Discussion Papers

Evaluating the time-varying impact of economic data on the accuracy of stock market volatility forecasts

Annika Lindblad
University of Helsinki and HECER

Discussion Paper No. 430
June 2018

ISSN 1795-0562

Evaluating the time-varying impact of economic data on the accuracy of stock market volatility forecasts*

Abstract

I assess the time-variation in predictive ability arising from the inclusion of macroeconomic and financial data in a GARCH-MIDAS model for stock market volatility. I consider whether the relative forecasting performance is affected by the state of the business cycle or the market environment. Results suggest predictive ability varies significantly over time, especially over long horizons. A central result is that models including macroeconomic data are useful for forecasting in low volatility periods. On the other hand, financial data performs overall surprisingly poorly. No single forecasting model or combination scheme is superior on all horizons and in all time periods, and while the term spread improves forecasting performance in particular over long horizons, forecast combinations perform well over the medium term.

JEL Classification: G17, E44, C52

Keywords: stock market volatility, MIDAS, volatility forecasting

Annika Lindblad

Economics
P.O. Box 17 (Arkadiankatu 7)
FI-00014 University of Helsinki
FINLAND

e-mail: annika.lindblad@helsinki.fi

* I thank Markku Lanne, Henri Nyberg and Charlotte Christiansen, as well as the participants of the FDPE Econometrics Workshops, the CREATES lunch seminar, the 9th Nordic Econometric Meeting and the 15th INFINITI Conference on International Finance for useful suggestions and comments. I gratefully acknowledge financial support from the OP Group Research Foundation, the Yrjö Jahnsson Foundation and the Academy of Finland.

1 Introduction

Forecasting volatility is a crucial part of decision-making for financial market actors as well as policy-makers. Long-horizon forecasts for volatility can be important for instance for portfolio allocation and risk management. While standard GARCH models are accurate for short-term return volatility forecasts (e.g., Andersen and Bollerslev (1998)), using models which also include economic data, such as the GARCH-MIDAS model, have been found to be successful at longer horizons (e.g., Engle et al. (2013), Conrad and Loch (2014)). There is mounting evidence that forecast accuracy varies over time (e.g., Giacomini and Rossi (2010) and Stock and Watson (2003)), and that predictability varies over economic states (e.g., Chauvet and Potter (2013), and for stock returns Rapach et al. (2010)). When using GARCH-MIDAS models for US stock market volatility we can see that the in-sample explanatory power of economic variables varies over time.¹ In particular, the ability of many macroeconomic variables to explain stock return volatility declines over time, which motivates studying the time-variation in forecasting performance of GARCH-MIDAS models driven by economic data. However, the stock market volatility forecasting literature largely concentrates on average forecasting performance over the whole out-of-sample period.

This paper explores the additional time-varying predictive ability provided by macroeconomic and financial variables using US data, by comparing the evolution of the out-of-sample forecasting performance of GARCH-MIDAS models to a standard asymmetric GARCH model. To consider potential reasons for the time-variation I investigate whether the relative forecasting performance is affected by the state of the business cycle or the market environment. While focus is on the out-of-sample analysis, also in-sample results are interesting because financial data have not been separately incorporated into a GARCH-MIDAS framework before². Finally, I determine whether forecast accuracy can be improved by combining the individual GARCH-MIDAS model forecasts, taking advantage of the (potential) detected time-variation. Focus is thus on improving real-time forecasts of long-term stock market volatility, with the data set representing as far as possible the information set of the forecaster at the forecast origin. The paper thus contributes to the current literature in three ways. First, it establishes the time-variation in the additional predictive ability provided by macroeconomic and financial variables in a GARCH-MIDAS context. Second, it considers whether the time-variation is related to different economic

¹See Figure 1 in Section 5.

²Asgharian et al. (2013) included financial data through principal components in a GARCH-MIDAS model.

or market environments. Lastly, the paper evaluates the performance of forecast combinations of GARCH-MIDAS model forecasts.

My results suggest that when forecasting over long horizons there are clear shifts in forecasting performance over time implying that (time-varying) forecast combination methods could be useful. Macroeconomic variables – in particular housing starts – improve predictions especially in low volatility periods but also in periods of weak economic growth, while financial data struggles to identify a long-term component in volatility, leading to weak forecasting performance. However, although some forecast errors are predictable conditioning on especially the volatility environment, it is difficult to achieve significant improvements in forecast accuracy in real-time forecasting. It is clear that no single forecasting model or combination scheme performs well on all horizons and in all time periods. When forecasting 12 months ahead the best forecasting model is the term spread driven GARCH-MIDAS model, while when forecasting 3 or 6 months ahead you would most likely choose a forecast combination method. Over the 1 month horizon there is some evidence that a GARCH-MIDAS model, or a combination method, currently performs best, although there are no statistically significant differences and the GJR-GARCH model performed well in the first half of the sample. The GJR-GARCH model is rarely significantly better than the GARCH-MIDAS models and never significantly outperforms the combination forecasts, indicating economic data is useful for long horizon forecasts.

The paper is organised as follows. Section 2 discusses the relevant literature, while Section 3 presents the GARCH-MIDAS model and the forecasting set-up. The data set is introduced in Section 4, and Section 5 briefly establishes in-sample results. When discussing the out-of-sample results in Section 6, I first present baseline full-sample results, before looking into the time-variation in forecasting performance. I consider forecast combination methods in Section 7, before concluding in Section 8.

2 Literature review

When forecasting stock return volatility focus has been on one-period-ahead forecasts where the step tends to be relatively short. For example, Engle (1982), Bollerslev (1986), Andersen and Bollerslev (1998) and Hansen and Lunde (2005) all considered one-step-ahead forecasts. Over short horizons, standard GARCH(1,1) models usually perform well. Poon and Granger (2003) thoroughly reviewed the volatility forecasting literature, and I will here concentrate on

the literature considering long horizon forecasts and models incorporating economic data (i.e., exogenous predictors). Ghysels et al. (2009) discussed multi-horizon volatility forecasts, comparing iterated, direct and MIXed DATA Sampling (MIDAS) approaches to the commonly used rule-of-thumb, where volatility is scaled up by the number of trading days. They found that for long horizons (over 30 days ahead) the MIDAS regression forecasts dominate. Ghysels et al. (2009) thus argued that volatility is in fact forecastable also over long horizons, contrary to the evidence in Christoffersen and Diebold (2000). Their study does not, however, consider GARCH-MIDAS models, or include macro-finance variables to enhance volatility forecasts.

There is ample evidence that stock return volatility is higher in recessions than in expansions (e.g., Schwert (1989)). Nevertheless, mixed results on the usefulness of economic data for modelling and forecasting volatility is found in, for example, Davis and Kutan (2003), Errunza and Hogan (1998), Pierdzioch et al. (2008) and Paye (2012). Other papers, such as Hamilton and Lin (1996), Cakmakli and van Dijk (2010), Christiansen et al. (2012) and Diebold and Yilmaz (2008) were more successful in linking economic developments to return volatility. These papers mostly rely on predictive regressions and VARs. Papers building on the component GARCH framework, introduced by Engle and Lee (1999), have successfully linked macroeconomic variables and stock market volatility. In particular, Engle et al. (2013) introduced the GARCH-MIDAS model, which decomposes volatility into a short-term component that fluctuates around a long-term trend determined by economic data. Conrad and Loch (2014), Asgharian et al. (2013), Asgharian et al. (2015) and Lindblad (2017) used the GARCH-MIDAS model to show that economic data helps explain and forecast stock return volatility.

Following the literature on time-variation in the accuracy of macroeconomic (e.g., Stock and Watson (2003)) and stock return (Rapach et al., 2010) forecasts, it is natural to think that the ability of economic data to forecast return volatility could be time-varying and depend on the state of the business cycle or the volatility environment. Christiansen et al. (2012), using predictive regressions and a Bayesian Model Averaging approach, compared the dynamic out-of-sample performance to autoregressive benchmarks, concluding that macro-finance variables add to predictability over the most recent financial crisis period. Paye (2012) found using predictive regressions that macroeconomic variables are especially useful for forecasting volatility around recessions, while Conrad and Loch (2014) noted that models using macroeconomic data lead to better forecasts than the GARCH-MIDAS model driven by realised volatility between the past two recessions and since the beginning of the financial crisis.

3 Methodology

3.1 The GARCH-MIDAS model

The GARCH-MIDAS model by Engle et al. (2013) is a multiplicative two-component model for the conditional variance, where the high-frequency component is modelled as a standard GARCH process, while the low-frequency component is determined by economic data.³ The high-frequency component can be thought of as fluctuating around a slow-moving long-term trend, which is driven by variables evolving at a lower frequency than returns. The MIXed DATA Sampling (MIDAS) approach, introduced by Ghysels et al. (2004)⁴, deals with the challenges related to using data sampled at different frequencies within the same model. The key feature of MIDAS is capturing the lag structure of the explanatory variables by a known function which depends on only a few parameters.

Following the interpretation in Engle and Rangel (2008), which builds on the log-linear dividend-ratio model in Campbell (1991) and Campbell and Shiller (1988), the stock return on day i and in period (month or quarter) t can be modelled as having a multiplicative specification for the conditional variance:

$$r_{i,t} = E_{i-1,t}(r_{i,t}) + \sqrt{\tau_{i,t} g_{i,t}} \varepsilon_{i,t}, \quad \varepsilon_{i,t} \mid \Phi_{i-1,t} \sim N(0, 1), \quad \forall i = 1, \dots, N_t$$

where $\Phi_{i-1,t}$ represents the information set up to day $i - 1$, and N_t is the number of trading days in period t . $\sigma_{i,t}^2 = \tau_{i,t} g_{i,t}$ is the total conditional variance, where τ_t ⁵ is the long-term volatility component and $g_{i,t}$ the GARCH component. The expected return is assumed constant: $E_{i-1,t}(r_{i,t}) = \mu$.

It is well established that stock return volatility is asymmetric (see e.g. Awartani and Corradi (2005) and the references therein), i.e., that positive and negative news have different impact on volatility. Therefore I use the asymmetric GJR-GARCH model (by Glosten et al. (1993)):

$$g_{i,t} = \omega + (\alpha + \gamma D_{i-1,t}) \frac{(r_{i-1,t} - \mu)^2}{\tau_t} + \beta g_{i-1,t} \quad (1)$$

where $D_{i-1,t}$ is an indicator function, taking the value 1 when $(r_{i-1,t} - \mu) < 0$ and 0 otherwise. Thus, γ describes the degree of asymmetry in volatility. ω is normalised to $\omega = 1 - \alpha - \beta - \gamma/2$

³The presentation of the model follows closely Engle et al. (2013).

⁴Discussed in detail in Ghysels et al. (2004), Ghysels et al. (2005), Ghysels et al. (2006), Ghysels et al. (2007), Andreou et al. (2010), and Wang and Ghysels (2015).

⁵ $\tau_{i,t}$ is fixed for all i in period t , so I drop the subscript i to ease notation and emphasise that τ_t evolves at a lower frequency than $g_{i,t}$.

so that $E(g_{i,t}) = 1$. To ensure stationarity the condition $\alpha + \beta + \gamma/2 < 1$ is imposed. In addition, I assume $\alpha > 0$, $\beta \geq 0$ and $\alpha + \gamma \geq 0$ to ensure the variance remains positive.

Following Engle et al. (2013) the MIDAS polynomial with one explanatory variable (X , which is, for example, a macroeconomic variable) takes the form:

$$\log \tau_t = m + \theta \sum_{k=1}^K \varphi_k(\omega_1, \omega_2) X_{t-k} \quad (2)$$

where $\varphi_k(\omega_1, \omega_2)$ is a weighting scheme, and K is the number of lags of explanatory data included. The logarithmic specification ensures non-negativity of the long-term volatility component (τ_t) even when the explanatory variable takes negative values. If the variable does not affect stock market volatility (i.e., $\theta = 0$), all volatility is captured by the short-term component and the model collapses to the GJR-GARCH model with $\tau_t = m$, i.e., unconditional volatility is constant. The standard GARCH model is therefore nested in the GARCH-MIDAS specification. The sign of θ is interpretable: $\theta > 0$ ($\theta < 0$) implies that higher values of X are linked to higher (lower) long-term volatility in stock returns.

A commonly used flexible but parsimonious weighting scheme is the beta lag polynomial⁶, which guarantees positive weights (which ensures non-negativity of volatility) that add up to one (this normalisation allows identifying θ):

$$\varphi_k(\omega_1, \omega_2) = \frac{(\frac{k}{K})^{\omega_1-1} (1-\frac{k}{K})^{\omega_2-1}}{\sum_{j=1}^K (\frac{j}{K})^{\omega_1-1} (1-\frac{j}{K})^{\omega_2-1}}, \quad \text{where } \sum_{k=1}^K \varphi_k(\omega_1, \omega_2) = 1.$$

The weight parameters, ω_1 and ω_2 , govern the shape of the weighting scheme and can be estimated or fixed before estimation. The beta polynomial allows both monotonously decreasing weights ($\omega_1 = 1$) and hump-shaped weights ($\omega_1 < \omega_2$). If $\omega_1 = 1$ the rate of decay is determined by ω_2 , where a larger value indicates faster decay. If $\omega_2 < \omega_1$ all weight can be on distant lags, which can be seen as counterintuitive. If $\omega_1 = \omega_2 = 1$ the weights are equal ($1/K$) for all lags, which corresponds to a moving average.

To assess how much the variation in a particular variable explains of the overall expected volatility, Engle et al. (2013) suggested calculating variance ratios: $\frac{Var(\log(\tau_t))}{Var(\log(\tau_t g_{i,t}))}$. The variance ratio can be interpreted as a measure of fit in the sense that the higher the variance ratio is, the larger is the share of the total expected volatility that can be explained by the variation in the long-term component. The GARCH-MIDAS model can be estimated using maximum likelihood

⁶Weighting schemes are discussed in more detail in Ghysels et al. (2007).

(or quasi-maximum likelihood if the assumption of normally distributed errors does not hold).⁷

3.2 Forecasting with the GARCH-MIDAS model

The one-step ahead volatility prediction is given directly by equations 1 and 2. For further horizons we iterate forward the daily GJR-GARCH model forecasts and combine this short-term forecast with a forecast for the long-term component, τ_t . For the GJR-GARCH model the forecast for day i is formed as:

$$E[g_{i,t}|F_{N_{t-1},t-1}] = 1 + (\alpha + \beta + \gamma/2)^{i-1}(g_{1,t} - 1) \quad (3)$$

where N_t is the number of trading days in period t , and $F_{N_{t-1},t-1}$ denotes the information set in period $t - 1$. The forecast for total volatility for period t can be expressed as:

$$E\left[\sum_{i=1}^{N_t} g_{i,t} \tau_t \varepsilon_{i,t}^2 | F_{N_{t-1},t-1}\right] = \tau_t \left[N_t + (g_{1,t} - 1) \frac{1 - (\alpha + \beta + \gamma/2)^{N_t}}{1 - \alpha - \beta - \gamma/2} \right]. \quad (4)$$

Following Conrad and Loch (2014) I create non-overlapping monthly forecasts by summing the daily forecasts over the respective month while keeping τ_t fixed at its one-step ahead prediction for all horizons. Because the forecast of the GARCH component converges to its (constant) unconditional expectation as the forecast horizon increases, in the long run the forecast differences are entirely driven by the long-term components (τ_t).

3.3 Forecasting set-up

The GARCH-MIDAS model has relatively many parameters to estimate, meaning that the estimation period needs to be long enough. However, in order to detect time-variation in the out-of-sample forecasts the evaluation period needs to be long enough as well. I thus divide the whole sample (January 1973 - June 2017) roughly into half: the first estimation period is January 1973 - December 1994, and the out-of-sample evaluation period is January 1996 - June 2017. As the short-term GARCH components are similar across all GARCH-MIDAS specifications, the largest gains in forecasting from including economic variables is expected to be achieved over long horizons. I therefore consider forecast horizons from 1 to 12 months.

⁷While consistency and asymptotic normality of the QML estimator for the rolling window GARCH-MIDAS model with realised volatility was established in Wang and Ghysels (2015), it has not been shown for the more general GARCH-MIDAS model with macroeconomic variables.

For the out-of-sample evaluation I use a rolling window estimation scheme, i.e., the estimation window is shifted forward by one period and the model is re-estimated before the next set of forecasts is calculated. A rolling window estimation scheme takes into account potential parameter instability, which is important if the relationship between long-term stock market volatility and the economic variables changes over time. In addition, the forecast comparison methods used require that limited memory estimators are used.

The forecasts are evaluated against realised volatility calculated as the monthly sum of squared daily returns ($RV_t = \sum_{i=1}^{N_t} r_{i,t}^2$). Forecast accuracy of a model is measured as the absolute value of the forecast error. The widely used squared forecast errors put significant weight on the largest forecast errors, which is useful if one wants to emphasise large forecast errors over smaller ones. However, since I wish to study general forecasting performance over time, and not in particular during for example the financial crisis, I use absolute forecast errors. In addition, Poon and Granger (2003) note that when using squared returns as the quantity of interest and using squared errors as the measure of forecast accuracy, one is effectively comparing the fourth moments of the data, which can complicate the comparison. However, Patton (2011) argues that while the mean squared forecast error (MSFE) loss function is robust in the sense that using a noisy proxy for volatility (such as the sum of squared daily returns) does not change the ranking of forecasting models, the mean absolute forecast error (MAFE) loss function is not. This concern needs to be taken seriously, and therefore, as a robustness check, I report MSFE ratios in the appendix and discuss them where relevant. In general, all the results are quantitatively similar, but as expected statistical significance is weaker when using squared forecast errors.

The natural benchmark model is the GJR-GARCH(1,1) model, since it is nested in the GARCH-MIDAS specification. Using the GJR-GARCH(1,1) model as benchmark thus reveals whether economic variables are useful for forecasting stock return volatility.

3.4 Measuring the time-variation in forecasting performance

The accuracy of the forecasting framework is important, but there is often considerable uncertainty regarding the choice of model. Thus it is important to be able to test the relative forecasting performance of competing models, and to this end several frameworks have been developed.⁸ However, the relative forecasting performance of models might be time-varying due

⁸For example, Diebold and Mariano (1995), West (1996), McCracken (2000), Clark and McCracken (2001), Clark and West (2006) and Giacomini and White (2006).

to, for example, structural instability (Giacomini and Rossi, 2010). Whether the relative forecasting performance of two models has shifted over time is an interesting and important question to complement full-sample results. To this end Giacomini and Rossi (2010) proposed the Fluctuation test, where the idea is to compare scaled and centred h -step-ahead out-of-sample forecast losses calculated over rolling windows of size m :

$$F_{t,m} = \hat{\sigma}^{-1} m^{1/2} \sum_{j=t-m/2}^{t+m/2-1} \Delta L_j(\hat{a}_{1,j-h,R}, \hat{a}_{2,j-h,R}), \quad (5)$$

where $t = R + h + m/2, \dots, T - m/2 + 1$, R is the in-sample size, ΔL_j is the difference in two loss functions in period j , $\hat{\sigma}^2$ is a HAC estimator of the variance (σ^2) and \hat{a}_1 and \hat{a}_2 are the in-sample parameter estimates of each model.⁹ The Fluctuation test tests the null hypothesis that the local relative forecasting performance equals zero at each point in time: $H_0 : E[\Delta L_t(\hat{a}_{1,t-h,R}, \hat{a}_{2,t-h,R})] = 0$. The testing framework allows both nested and non-nested models as well as non-linear models, but the parameters need to be estimated using a limited memory estimation scheme, such as rolling windows. Giacomini and Rossi (2010) showed that if the ratio between m and $T - R$ (out-of-sample size) is too small, the Fluctuation test is oversized. The size of the test is found to be largely correct for $\frac{m}{T-R} \approx 0.3$. As my out-of-sample size is 258 I need, for example, $m = 78$, which corresponds to 6.5 years of monthly data. The test is therefore designed to detect long-term shifts in forecasting performance.

4 Data

I use the continuously compounded daily stock market return on the CRSP index from January 1973 to June 2017. From a theoretical perspective time-variation in stock return volatility can be linked to uncertainty regarding future cash flows, which can stem from, for example, uncertainty regarding the true macroeconomic situation and expectations regarding the future economic environment. As exogenous variables I include a collection of commonly used (monthly) predictors for stock return volatility, representing the financial markets, the macroeconomy and expectations regarding the economic environment. While the important role of many macroeconomic variables in driving long-term volatility has been established in the GARCH-MIDAS literature (see Section 2), financial variables have not been explored in the GARCH-MIDAS context before (with the exception of the term spread and realised volatility). Asgharian et al. (2013) included

⁹See Giacomini and Rossi (2010) for details.

the 3-month T-bill rate and a default spread, but they were aggregated together with macroeconomic variables into principal components. Using predictive regressions financial variables have been identified as important predictors of stock return volatility (e.g., Christiansen et al. (2012)).

The macroeconomic variables included are real-time housing starts (change in level), the real-time Aruoba-Diebold-Scotti Business Conditions index (ADS index)¹⁰, the Buying Conditions index (forward-looking sub-index of the University of Michigan consumer confidence index, change in level), and the ISM New Orders index (level). As a forward-looking indicator housing starts has been among the best predictors for stock return volatility (e.g., Conrad and Loch (2014)), the ADS index reflects the current economic situation, and the Buying Conditions index and the ISM New Orders index represent expectations of the macroeconomic situation.

As financial data I include commonly used equity return predictors (as in, e.g., Goyal and Welch (2008)), and in particular those found most useful for predicting stock return volatility in, e.g., Christiansen et al. (2012) and Conrad and Loch (2014).¹¹ I use a realised volatility measure (sum of the absolute value of daily returns: $RV_t = \sum_{i=1}^{N_t} |r_{i,t}|$), and the term spread (difference between the 10-year Treasury bond yield and the 3-month T-bill rate). In addition, I include the short term and long term interest rates (level and change over month), as well as the default spread (default risk of corporate bonds, difference between BAA and AAA bond yields), which describes credit risk. To capture equity market movements I include excess market returns. For missing values I use the previous month's data.¹² See Appendix A.1 for data sources.

To determine whether a broad set of macroeconomic and financial variables is useful for forecasting stock market volatility I use the dataset and methodology in McCracken and Ng (2016) to extract factors using principal components analysis. The dataset currently comprises 128 macroeconomic and financial variables. I use the first four principal components (PC) in the analysis, which explain a combined 34% of the total variation in the data.¹³ As shown in more detail in Appendix A.2, the first PC relates to real activity and employment, the second one concentrates on price variables, the third one relates mainly to interest rate spreads, while

¹⁰Includes, for example, industrial production and labour market data, see <https://www.philadelphiafed.org/research-and-data/real-time-center/business-conditions-index> for details. Prior to 2008 real-time vintages are not available.

¹¹A requirement is that data is available from January 1971 until June 2017 (up to two years of economic data is needed to estimate the model for the first period). Therefore, for example, the investor sentiment index by Baker and Wurgler (2006) (available until September 2015) and the E/P and D/P ratios (available until December 2016) are not included, although they have been successful predictors for returns. In results which are available upon request I determine that these variables are not important drivers of long-term stock market volatility.

¹²This is important for the Buying conditions index, which is available at a quarterly frequency before 1978.

¹³See McCracken and Ng (2016) for details on the data, the extracted factors (which are very similar to those extracted here) and the methodology.

the fourth one is dominated by financial variables. I use as far as possible real-time data for the principal components in the rolling window analysis. Historical vintages go back to August 1999. Before that I use the August 1999 vintage and recursively estimate the PCs for each period, so that only historical data is used.¹⁴ The first time-varying PC relates mainly to the same underlying macro series – real activity and employment related series – as the full-sample PC, as shown in Figure A.2. For the second and third PCs the compositions vary, although interpretation of the factors remains relatively constant over time. The second PC mainly relates to interest rates and interest rate spreads, but also to price variables, as in the full-sample results. For the third PC one cluster relates to price variables, a second to interest rates, and a third one relates to housing market data.

5 In-sample results

First, I establish in-sample results for the full-sample period, then, I look at parameter stability over the forecasting horizon using a rolling window estimation scheme. Importantly, it will reveal how the long-term relationship between economic variables and stock market volatility has changed over time, as identified by the GARCH-MIDAS model.

5.1 Full-sample results

In the MIDAS polynomial lag length K needs to be determined. I choose between $K = 12$ and $K = 24$ for each model, i.e., one or two years of lagged economic data, and proceed with the lag length maximising the log-likelihood function value.¹⁵ The same K is used throughout the rolling window estimations.

Table 1 presents the in-sample estimation results over the full sample of all the GARCH-MIDAS models and the baseline GJR-GARCH model. The macroeconomic data, the term spread and realised volatility get highly significant estimates for θ as well as high variance ratios, implying the variables are useful for modelling stock market volatility.¹⁶ These results largely echo earlier results (Conrad and Loch (2014) and Lindblad (2017)).

¹⁴Note that the number of series in the data set varies between 106 and 135. See Appendix A.3 for details.

¹⁵The results are not, however, materially changed by the choice of 12 or 24 lags.

¹⁶Notice that when testing the significance of θ , θ and the weight parameters ω_i are not separately identified under the null hypothesis, which affects the asymptotic distribution of the test statistic. However, I follow the convention in the GARCH-MIDAS literature (e.g., Engle et al. (2013), Conrad and Loch (2014)) and proceed using the standard t-statistic. In addition, Appendix A.6 discusses estimates of θ using a fixed weighting scheme. See Ghysels et al. (2007) for a discussion of the problem in MIDAS regressions.

Table 1: Estimation results for GARCH-MIDAS model with one explanatory variable

	μ	α	β	γ	θ	ω_1	ω_2	m	VR	LLF	K
GJR-GARCH(1,1)	0.0466*** (0.0074)	0.0217*** (0.0050)	0.9024*** (0.0136)	0.1073*** (0.0180)	-	-	-	0.8500*** (0.0872)	-	-14281.88	-
Realised volatility	0.0482*** (0.0073)	0.0133** (0.0056)	0.8559*** (0.0175)	0.1438*** (0.0203)	0.0639*** (0.0050)	2.5509*** (0.9418)	6.5085** (2.6167)	-1.1786*** (0.0979)	34.79	-14229.19 [0.0094]	12
Buying Conditions	0.0454*** (0.0074)	0.0182*** (0.0052)	0.8936*** (0.0143)	0.1174*** (0.0186)	-0.1788*** (0.0259)	1.8624*** (0.4391)	2.1397*** (0.7643)	-0.1588* (0.0844)	14.24	-14253.96 [0.0010]	24
ISM New Orders	0.0456*** (0.0074)	0.0144*** (0.0054)	0.8987*** (0.0138)	0.1188*** (0.0183)	-0.0522*** (0.0086)	1	2.6036*** (0.9094)	2.6681*** (0.4760)	15.37	-14254.51 [0.6112]	24
ADS index	0.0464*** (0.0074)	0.0159*** (0.0053)	0.8968*** (0.0138)	0.1174*** (0.0183)	-0.4817*** (0.0761)	1	3.3587*** (0.8423)	-0.2496*** (0.0874)	15.09	-14255.13 [0.6067]	24
Housing starts	0.0463*** (0.0074)	0.0170*** (0.0052)	0.8952*** (0.0143)	0.1179*** (0.0185)	-0.0150*** (0.0022)	2.0944*** (0.6512)	1.7774*** (0.4682)	-0.2137** (0.0851)	17.59	-14249.16 [0.0000]	24
Term spread	0.0468*** (0.0073)	0.0174*** (0.0052)	0.8933*** (0.0149)	0.1174*** (0.0192)	-0.2485*** (0.0417)	2.8814 (2.5458)	1.6183* (0.8912)	0.2411** (0.1095)	13.87	-14255.26 [0.0148]	24
Default spread	0.0456*** (0.0073)	0.0133*** (0.0051)	0.8977*** (0.0144)	0.1217*** (0.0193)	0.5605*** (0.0994)	1	6.7455** (2.9512)	-0.8116*** (0.1500)	12.07	-14261.93 [0.2594]	12
3M T-bill rate (level)	0.0456*** (0.0073)	0.0177*** (0.0052)	0.9028*** (0.0139)	0.1127*** (0.0187)	0.0437*** (0.0157)	300 (499.2185)	233.5683 (402.5675)	-0.3906*** (0.1278)	4.46	-14273.72 [0.0441]	24
3M T-bill rate (chg over month)	0.0458*** (0.0074)	0.0175*** (0.0052)	0.9020*** (0.0135)	0.1126*** (0.0181)	-0.7768** (0.3249)	1	1.7220* (0.8999)	-0.1821 (0.0959)	3.18	-14275.57 [0.0795]	12
10Y Treasury rate (level)	0.0462*** (0.0073)	0.0203*** (0.0051)	0.9030*** (0.0137)	0.1090*** (0.0183)	0.0221 (0.0185)	1	1.0000 (3.7585)	-0.3145** (0.1601)	0.83	-14280.65 [0.3694]	24
10Y Treasury rate (chg over month)	0.0467*** (0.0074)	0.0204*** (0.0050)	0.9029*** (0.0132)	0.1090*** (0.0176)	-0.6228 (0.3525)	5.2828* (2.5327)	34.9021 (22.6232)	-0.1704 (0.1016)	2.20	-14275.41 [0.0032]	24
Excess market return	0.0479*** (0.0073)	0.0159*** (0.0048)	0.9066*** (0.0112)	0.1165*** (0.0157)	0.1089*** (0.0286)	1	3.8440*** (0.8528)	-0.2337* (0.1135)	9.30	-14262.94 [1.0000]	12
Principal component 1	0.0466*** (0.0074)	0.0163*** (0.0053)	0.8944*** (0.0141)	0.1194*** (0.0185)	0.9380*** (0.1450)	1	6.9868** (2.9702)	-0.2252*** (0.0847)	16.17	-14254.73 [0.6032]	24
Principal component 2	0.0459*** (0.0074)	0.0174*** (0.0053)	0.8970*** (0.0144)	0.1171*** (0.0194)	-1.8320*** (0.4827)	12.0342 (28.1765)	6.2691 (16.0309)	-0.1859** (0.0902)	10.01	-14263.10 [0.0216]	24
Principal component 3	0.0458*** (0.0074)	0.0172*** (0.0052)	0.8988*** (0.0145)	0.1154*** (0.0191)	1.0902*** (0.2500)	4.5780 (3.7068)	2.3205 (1.6009)	-0.1804** (0.0913)	10.55	-14263.59 [0.0031]	24
Principal component 4	0.0467*** (0.0074)	0.0210*** (0.0051)	0.9012*** (0.0137)	0.1089*** (0.0181)	-0.7049** (0.3452)	12.4661 (8.3094)	30.6975 (26.2808)	-0.1723* (0.0987)	2.47	-14276.94 [0.0027]	24

Bollerslev-Wooldridge QMLE robust standard errors are reported below the parameter estimates. *, ** and *** indicate significance at the 10%, 5%, and 1% level, respectively. VR is the variance ratio from Section 3.1, multiplied by 100. MIDAS polynomial: $\log \tau_t = m + \theta \sum_{k=1}^K \varphi_k(\omega_1, \omega_2) X_{t-k}$, where X stands for the explanatory data, stated in the first column. All models are estimated with a restricted ($\omega_1 = 1$) and an unrestricted weighting scheme. The model reported in the table is chosen based on a likelihood ratio test between the restricted and unrestricted specifications. The related p-value is reported below the value of the log likelihood function (LLF).

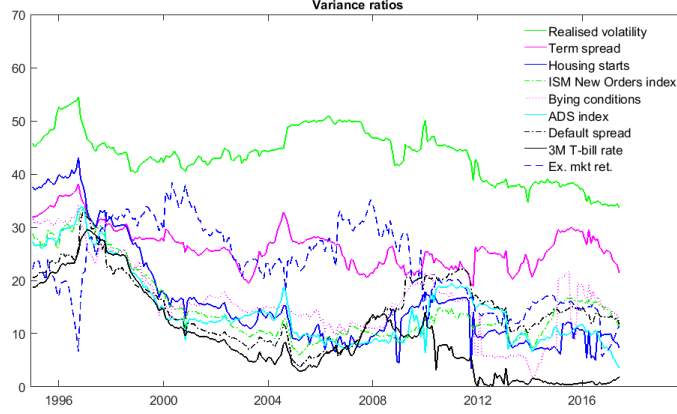


Figure 1: Variance ratios of selected GARCH-MIDAS models. Based on monthly rolling windows: first period January 1973-December 1994, last period July 1994-June 2016.

The interest rate data does not lead to good in-sample fit, as evidenced by both weakly significant parameter estimates and low variance ratios, and as such these models are unlikely to produce forecasts very different from the baseline GJR-GARCH(1,1) model. I therefore only include the 3M T-bill rate (level), which gets the highest variance ratio and a highly significant parameter estimate, in the subsequent out-of-sample analysis.¹⁷

The default spread, 3M T-bill rate, and excess market return get positive estimates for θ , implying that a higher risk of default, a higher interest rate and a higher excess market return lead to higher stock return volatility. The first PC explains a large, 16% share of the total variance, while the two following factors explain roughly 10% each. The estimates for θ are also highly significant. On the other hand, the fourth PC has a very low variance ratio, implying it is not an important driver of long-term volatility. I thus proceed using the first three factors.

Figure 1 shows how the in-sample explanatory power of various GARCH-MIDAS models varies over time, as indicated by the variance ratio calculated over rolling windows. The GARCH-MIDAS model where the long-term volatility component is driven by lagged realised volatility (RV) explains a stable 40%-50% of total volatility, while the long-term component of the model driven by the term spread explains a relatively stable 20%-30%. For the remaining models the explanatory power of the economic variables seems to decline over time.

¹⁷The weighting scheme of the 3M T-bill rate (level) is somewhat counterintuitive, with the parameter estimate for ω_1 reaching the upper bound of the parameter space. However, I am mostly interested in the rolling window estimates of the model parameters, discussed in Section 5.2.

5.2 Parameter instability

There are 270 out-of-sample months (January 1995 - June 2017), and hence 270 estimates for each parameter. In this section I discuss how the parameter estimates vary over time, and how representative the full sample results are. I will also examine whether the choice of restricted or unrestricted weighting scheme remains constant over the out-of-sample period.¹⁸ Taking into account parameter instability is important for forecasting if there are structural breaks, and thus the relationship between stock return volatility and the economic variables changes over time.

Regarding the choice between a restricted and unrestricted model, Table 2 presents the percentage of times the unrestricted weighting scheme is chosen over the restricted one, chosen by a likelihood ratio test in each of the 270 out-of-sample periods.¹⁹ Clearly, for realised volatility, the ISM New Orders index and PC1 the restricted model is always chosen, while for the Buying conditions index and housing starts we always choose the unrestricted weighting scheme. For the ADS index, the default spread and the 3M T-bill rate we almost always choose the restricted weighting scheme. On the other hand, for the term spread, the excess market return, PC2 and PC3 the choice varies, although the unrestricted weighting scheme is chosen more often.²⁰

Table 2: Choice between restricted and unrestricted weighting scheme

	% of total		% of total		% of total
Buying conditions index	100	ADS index	4.81	3M T-bill rate	7.41
ISM New Orders index	0	Term spread	54.44	Excess market return	69.63
Housing starts	100	Default spread	2.22	Realised volatility	0
PC 1	0	PC 2	65.56	PC 3	78.89

The table reports the percentage of times the unrestricted weighting scheme is chosen over the restricted one, i.e., if the number is over 50 the unrestricted weighting scheme is chosen more often than the restricted weighting scheme. The choice was made based on a likelihood ratio test (at the 5% confidence level) in each of the 270 out-of-sample periods. See Appendix A.4 for details.

Figure 2 plots the time-variation in the estimated GARCH parameters, as well as the time-variation in the statistical significance of γ , which describes the degree of asymmetry in returns. The parameters relating to the GARCH model behave very similarly over time and in line with the baseline GJR-GARCH(1,1) model. The exception is the GARCH-MIDAS model driven by realised volatility, for which especially β is estimated lower and γ higher compared to the other models. Interestingly γ roughly doubles in magnitude over time in all models. This implies that

¹⁸The graphs in this section as well as the out-of-sample analysis in Section 6 are based on the weighting scheme which is chosen more often. See robustness checks in Appendix A.4.

¹⁹Appendix A.4 discusses in more detail the time-variation in the choice of weighting scheme and the implications of choosing a particular weighting scheme. Overall the differences are relatively small.

²⁰Notice that the variation in the optimal weighting scheme for the principal components can also be a result of the changing composition of the PC itself.

smaller-than-expected returns (with estimated parameter $\alpha + \gamma$) affect volatility much more than larger-than-expected returns (with estimated parameter α), and this effect becomes more pronounced towards the end of the sample period. γ also remains significantly different from zero for all models in most periods (see Panel 2e).

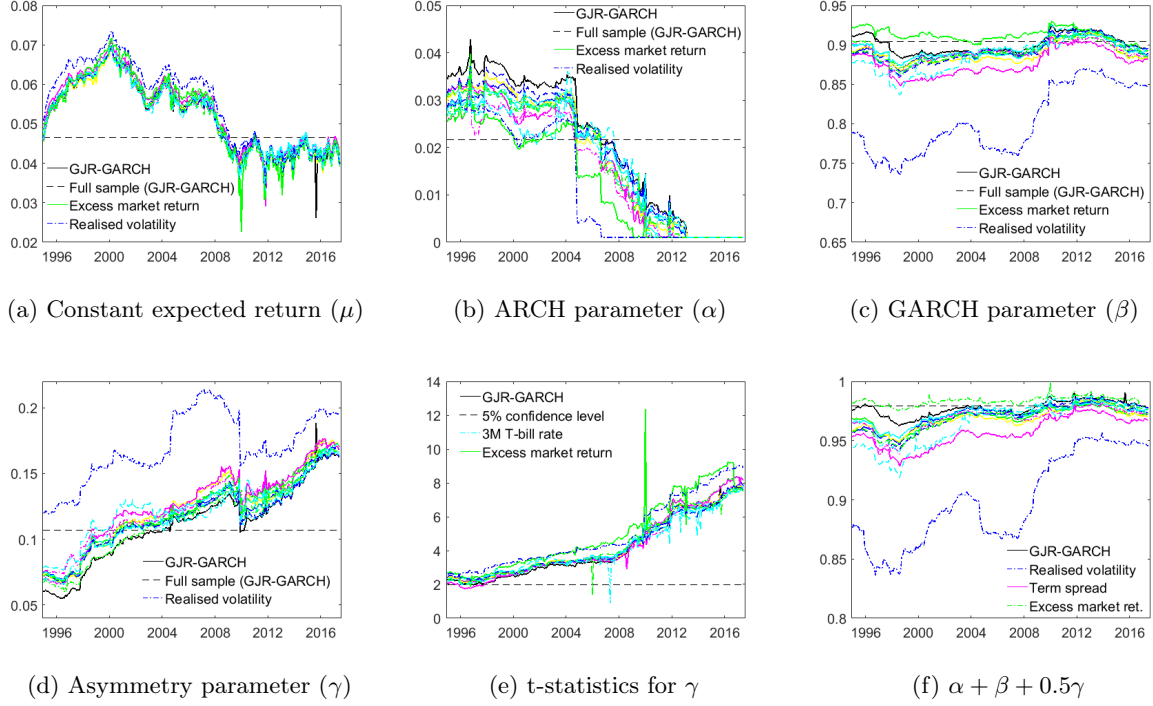


Figure 2: Time-variation in GJR-GARCH model parameters. Legends contain selected series.

The relationship between economic data and stock return volatility is described by θ . Figure 3 shows how the estimates for θ change over the out-of-sample period in the different GARCH-MIDAS specifications. Mostly θ fluctuates around the full-sample estimate, but, for example, for realised volatility there is a time trend in θ , indicating a rolling estimation scheme is appropriate. Counterintuitively the sign of θ for the excess market return changes at the end of the sample period. For the second and third principal components the sign of θ varies over time, resulting, most likely, from the time-varying correlation with the underlying economic variables. In many specifications θ is significantly different from zero in most periods, confirming that economic data is important for long-term volatility. The main exceptions are the second and third principal components and the 3M T-bill rate, for which θ is, especially recently, not significantly different from zero at the 5% level, implying we could instead use the basic GJR-GARCH model.

It is also interesting to consider how the weight parameter(s) in the different GARCH-MIDAS specifications change over time. Figure 4 depicts the time-variation in the estimated

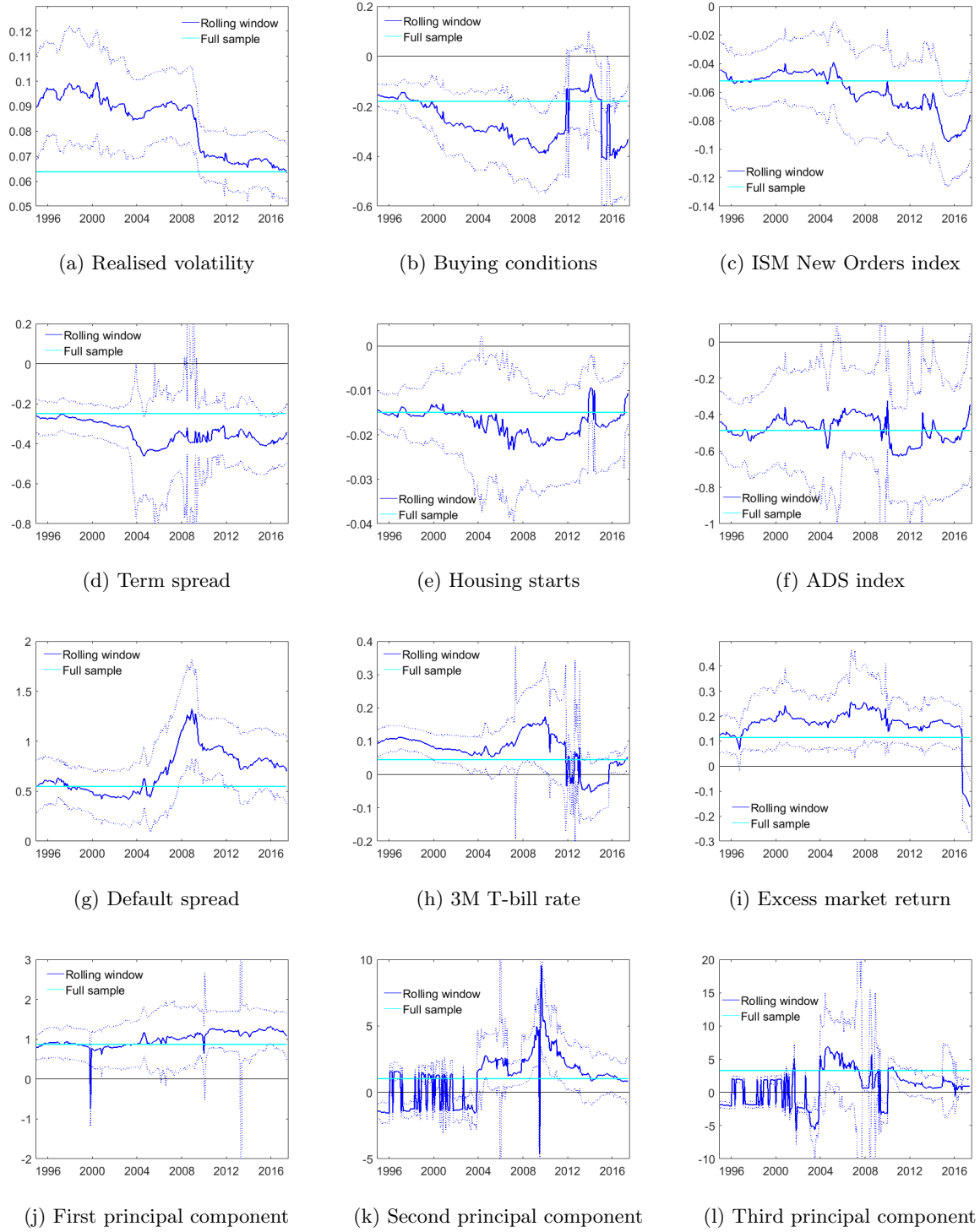


Figure 3: Time-variation in rolling window estimates of θ , compared to full-sample estimates. Dashed lines mark 5% confidence bands.

weight parameters for each of the GARCH-MIDAS models. The weight parameter (ω_2) for realised volatility and the ISM New Orders index is shrinking, implying that the decay of the weights becomes slower and further lags become increasingly important. The ADS index, the 3M T-bill rate, the default spread and the first PC mostly exhibit a similar weighting scheme

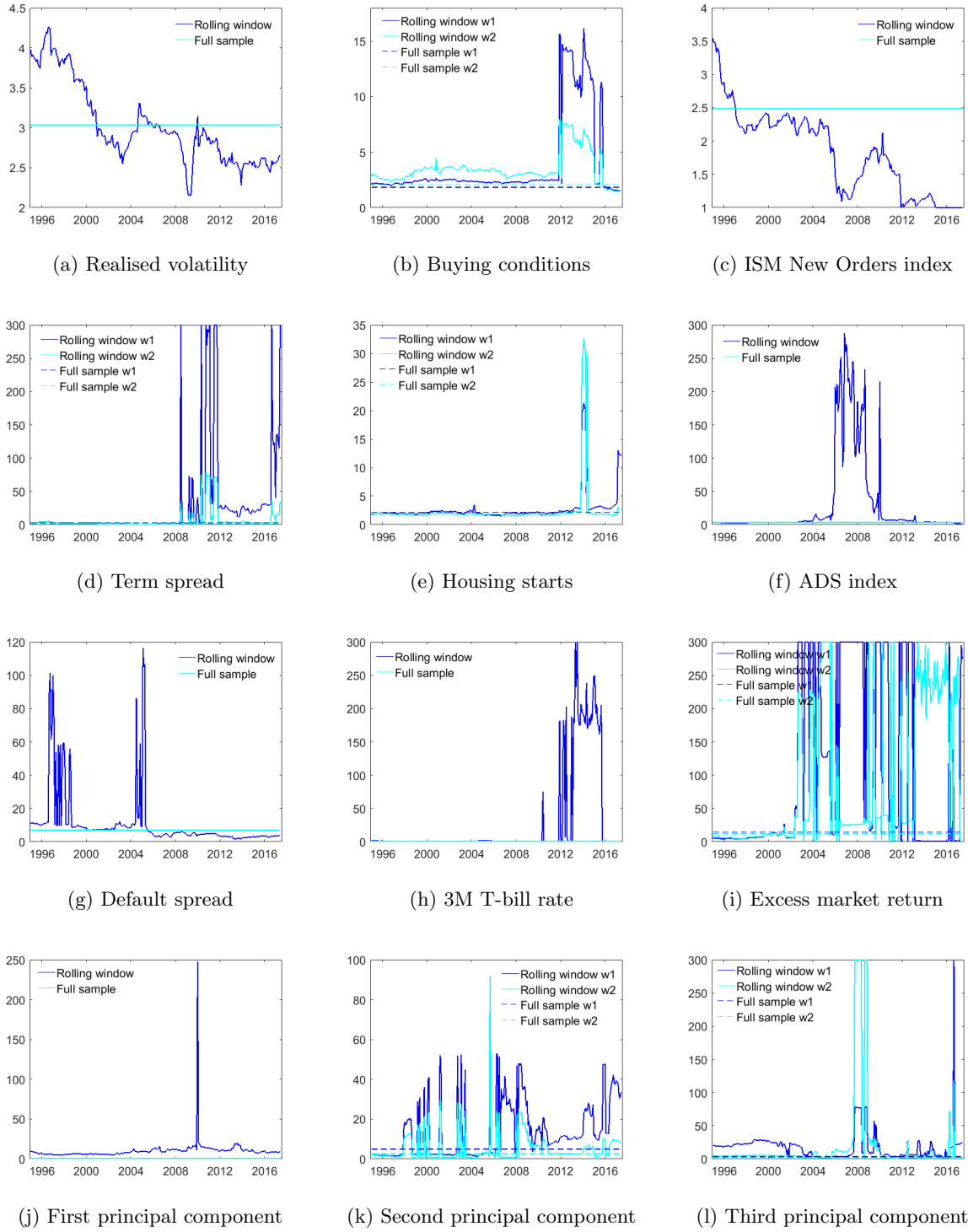


Figure 4: Time-variation in rolling window estimates of w , compared to full sample estimates.

to the full-sample results, but there are time periods when only the most recent data matters (i.e., ω_2 is very large). For the term spread and the Buying conditions index, and to a lesser degree housing starts, towards the end of the sample period there is a tendency for the weighting schemes to put significant weight on a specific lag, which is not necessarily the most recent one.

The time-variation in both the weighting schemes and the estimates for θ indicates that the relationship between economic data and long-term stock market volatility varies over time and that the chosen sample period matters. The variation in weights over time can reflect estimation problems (e.g., related to the small sample size), but can also be due to a changing relationship between the variables and volatility. This is of particular concern for the GARCH-MIDAS models driven by the excess market return, the term spread and the third PC, for which several of the weight parameters are imprecisely estimated and hit the upper bound (300) used in the estimation. To guard against estimation problems I re-estimate the models with weight parameters (ω_1 and ω_2) fixed at their full-sample values, as well as using an expanding window estimation scheme. The out-of-sample forecasting results remain largely robust (see Appendix A.6 for details).

6 Out-of-sample results

I first discuss the forecasting performance over the whole out-of-sample period, establishing a benchmark. Then, Section 6.2 looks at how the relative forecasting performance has changed over time, while Section 6.3 considers whether the forecasting performance varies with the economic environment. Section 6.4 discusses conditional predictive ability, i.e., whether the relative forecast losses are predictable using current information.

6.1 Forecasting performance over the whole out-of-sample period

The MAFE ratios in Table 3 indicate that the GJR-GARCH(1,1) model is hard to beat, at least in a statistically significant way.²¹ It can only be improved upon at longer horizons, and mainly by the GARCH-MIDAS model driven by the term spread or the second PC, which is in line with the forward-looking nature of these variables. Other financial variables fail to improve on the benchmark model at any horizon and in fact perform clearly worse in some cases. This is contrary to results using predictive regressions (see e.g. Christiansen et al. (2012)), and could reflect the fact that financial data fail to robustly extract a long-term trend of volatility, which is crucial for the GARCH-MIDAS model.²² The Mincer-Zarnowitz regressions in appendix A.11 show that mostly the forecasts are unbiased, although they become more biased as the forecast horizon increases.

²¹These results echo those in Lindblad (2017).

²²The MSFE ratios (Table A.3) convey a qualitatively similar picture.

Table 3: Full sample results: Mean absolute forecast error ratios

	1M ahead	2M ahead	3M ahead	6M ahead	9M ahead	12M ahead
Buying Conditions index	1.00	0.98	0.96*	0.96	0.97	0.98
ISM New Orders index	0.99	1.00	0.99	0.99	0.98	0.97
Housing starts	0.99	0.99	1.00	0.97	0.95	0.94
ADS index	1.03	1.00	1.00	0.99	1.00	1.00
Term spread	1.03	1.03	0.99	0.91***	0.88***	0.87***
Default spread	1.09	1.13	1.14	1.20	1.22	1.20
3M T-bill rate	1.01	1.02	1.02	1.02	1.02	1.01
Excess market return	1.08	1.03	1.02	1.04**	1.06**	1.08**
Realised volatility (RV)	1.14*	1.23**	1.22*	1.29	1.32	1.31
First principal component	0.99	1.00	1.00	1.02	1.04	1.02
Second principal component	1.01	0.99	0.98	0.96*	0.96	0.95**
Third principal component	1.05	1.00	0.99	0.98	0.98	0.99

Benchmark: GJR-GARCH(1,1). MAFE ratio: $\frac{MAFE_{GMX}}{MAFE_{GARCH}}$, where $MAFE_{GMX}$ stands for the mean absolute forecast error from the GARCH-MIDAS model driven by some macroeconomic or financial data (X). A value below 1 means the GARCH-MIDAS model outperforms the GJR-GARCH(1,1) model. *, ** and *** indicate a rejection of the null hypothesis of equal (unconditional) predictive ability at the 10%, 5%, and 1% level, respectively, according to the Giacomini and White (2006) test.

6.2 Time-varying forecasting performance

We saw in the previous section that many models forecast on average roughly equally well, for example, on the 3M horizon the MAFE ratio for housing starts and the ADS index equal one, and that of the ISM New Orders index and the term spread equal 0.99. However, this can either be because forecasting performance is similar across models in all time periods or there could be time-variation in relative performance which cancels out over time. To formally investigate time-variation in forecasting performance I use the Fluctuation test by Giacomini and Rossi (2010). Figure 5 plots the scaled difference in loss functions of a GARCH-MIDAS model and the GJR-GARCH(1,1) model (the test statistic, see equation 5), together with two-sided confidence bands.²³ For clarity I focus on a representative subset of the results, with the full results available in Appendix A.7. Each row corresponds to one economic variable, while the first column presents results for the 1 month forecasting horizon, the second column for the 3M horizon, the third one for the 6M horizon and the rightmost column for the 12M horizon. As the test statistic is calculated over a rolling 6.5 year period the first period is January 1996 - June 2002 and the last period covers January 2011 to June 2017. If the test statistic (solid blue line) exceeds the upper bound (dashed line) the GARCH-MIDAS model produces significantly worse forecasts

²³I set $\alpha = 0.1$ (significance level). I use a Newey-West estimator of the asymptotic variance matrix with lag length $l = 5$, based on the rule-of-thumb, $l = 0.75 \sqrt[1/3]{T} = 4.77$. The results are robust to changing the lag length to 4 or 8, results are available upon request. See Giacomini and Rossi (2010) for details on the test and the confidence bands.

than the baseline GJR-GARCH(1,1) model, if it drops below the lower bound (dashed line) then the loss of the benchmark model significantly exceeds the loss of the GARCH-MIDAS model. Generally, as long as the test statistic is negative the GARCH-MIDAS model outperforms the GJR-GARCH(1,1) model, and we can say that the explanatory variable is useful for forecasting volatility.

Overall, the forecast accuracy of the different GARCH-MIDAS models vary significantly over time, with the differences in performance becoming larger as the forecasting horizon increases. There is, however, no one model that is superior over all forecasting horizons. In general, the baseline GJR-GARCH model has only been significantly better than some of the GARCH-MIDAS models early in the sample period. Recently all the test statistics have been negative for all the GARCH-MIDAS models driven by macroeconomic data and the term spread, although only the GARCH-MIDAS model driven by housing starts has been able to outperform the baseline GJR-GARCH model in a statistically significant way (on the 6M horizon).²⁴ The benefit of augmenting a basic GARCH model with financial data (excluding the term spread) remains weak even at longer horizons. The GARCH-MIDAS model driven by the ISM New Orders index significantly outperforms the baseline model in at least one time period on most horizons, despite very small difference in full sample performance (see Table 3). On the 6M horizon the term spread and Buying conditions index driven models outperform the benchmark for the whole, or almost the whole, sample period, although the differences are mostly not statistically significant.

As expected, the statistically significant differences occur mainly at the 12 month horizon. In particular, the GARCH-MIDAS model driven by the term spread significantly outperforms the asymmetric GARCH(1,1) model for most of the sample period, and the test statistic remains close to the lower bound for the whole period. Thus it seems it is difficult to beat the term spread as a predictor for stock market volatility for any time period when forecasting 12 months ahead. For housing starts and some financial variables (such as the default spread) the GJR-GARCH(1,1) model first outperforms the GARCH-MIDAS models, but performance reverses so that mid-sample macroeconomic data is useful for forecasting. For housing starts the shifts are statistically significant, meaning there is significant time-variation in forecasting performance.

²⁴As expected, the results are weaker in terms of statistical significance when using mean squared forecasts errors, and the recent improvements in the relative performance of the GARCH-MIDAS models is less convincing, see Appendix A.5.

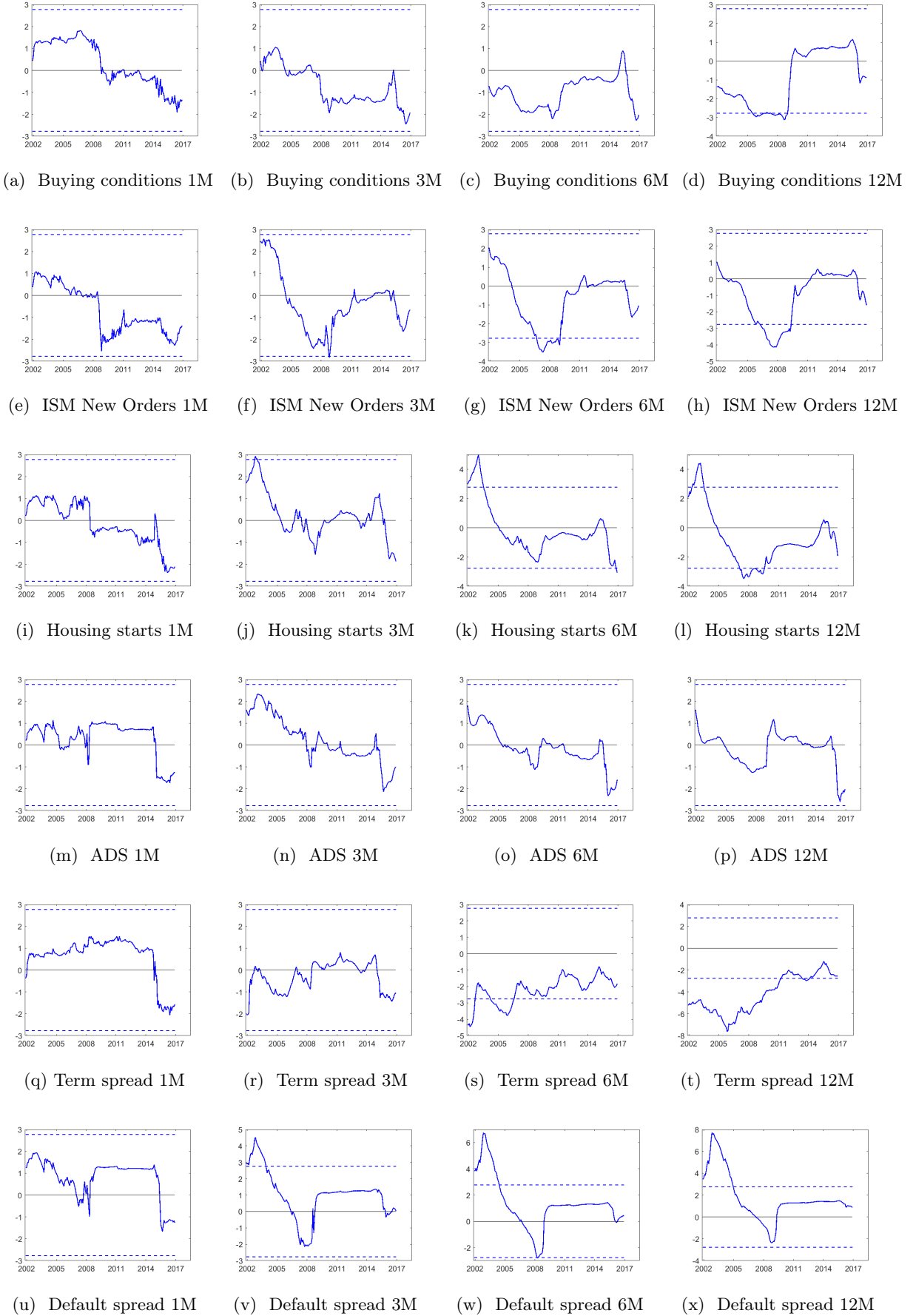


Figure 5: Fluctuation test results for loss function differences between the GARCH-MIDAS model driven by the economic data stated below the figure and the GJR-GARCH(1,1) model. Dashed lines represent 10% confidence bands. Note that the year on the x-axis marks the end of the rolling window period, over which the test statistic is calculated.

To see more specifically how the relative forecasting performance has evolved over time, Figure 6 plots the cumulative sum of loss function differences for three different forecast horizons: 1, 3 and 12 months ahead. For many GARCH-MIDAS models performance seems to significantly deteriorate during or immediately after the financial crisis in 2007-2008 (e.g., ADS index and default spread). On the other hand, housing starts are especially useful right before and during the latest recession (12 month horizon), while the good performance of the GARCH-MIDAS model driven by the ISM New Orders index documented earlier can be attributed to the ISM New Orders index improving forecasts especially between the two recessions on all horizons.

For the GARCH-MIDAS models driven by macroeconomic data and the term spread relative performance has improved since the latest recession, with the largest gains seen on the 12M horizon, while financial data has mainly performed no better than the GJR-GARCH model, as evidenced here by the default spread driven model. In addition, the Buying conditions index is a useful predictor also at the 3M horizon.

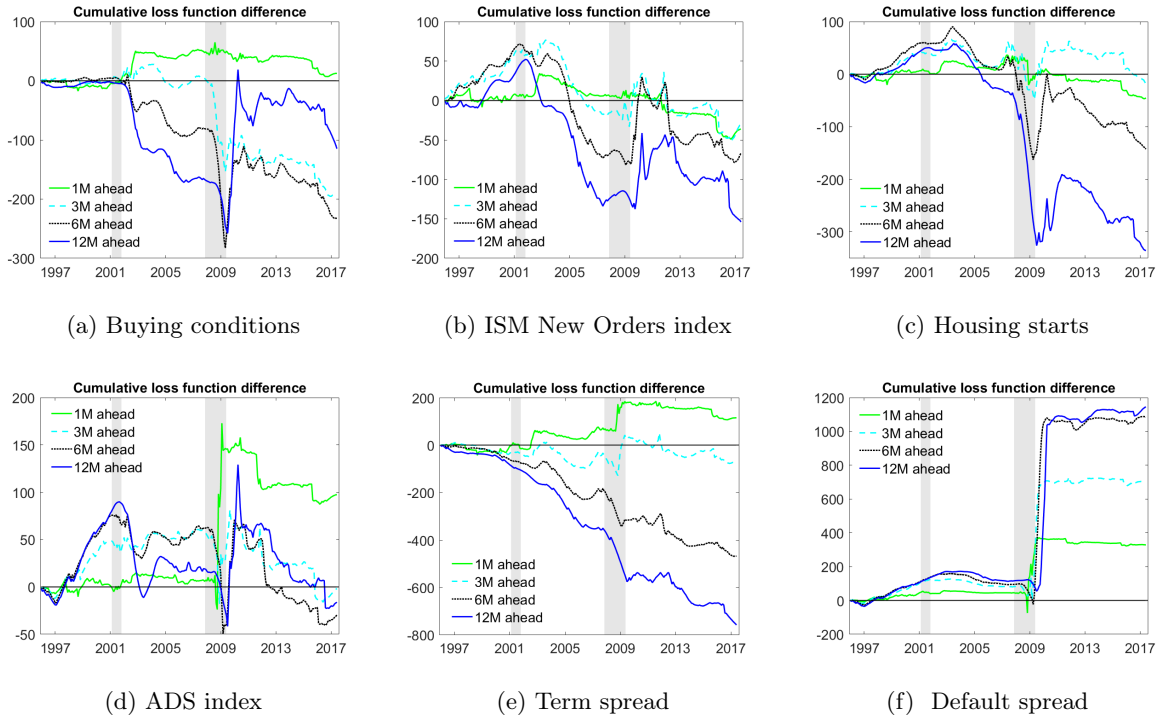


Figure 6: Cumulative sum of loss function differences (absolute errors) ($|Loss_{GMX}| - |Loss_{GARCH}|$). An upward sloping segment thus indicates the GJR-GARCH model outperforms the GARCH-MIDAS model. Grey areas mark NBER dated US recessions.

6.3 Effect of economic environment on forecasting performance

As shown above, the ability of economic data to predict long-term stock return volatility varies over time. However, is this purely random variation or can it be explained by the economic or market environment? As discussed in, for example, Hamilton and Lin (1996), it is logical to assume that the dynamic behaviour of the economy is different during expansions and contractions, and that the business cycle can thus be broken down into two distinct states. When forecasting volatility it is also plausible that the volatility environment can affect relative forecast accuracy.

I divide the out-of-sample period into sub-samples according to a business cycle (or volatility) indicator, and compare forecasting performance separately for recession (or high volatility) and expansion (or low volatility) periods. If we, for example, anticipate entering a recession (high volatility period) this approach can help us choose a more accurate forecasting model.

6.3.1 Business cycles

I first divide the sample into positive and negative growth periods based on the sign of industrial production growth. As a robustness check I divide the data based on the NBER dated US recessions (see Appendix A.8 for the robustness checks and plots of the regimes).

Table 4: Effect of business cycle (IP growth) on forecasting performance: MAFE ratios

	1 month ahead		3 month ahead		6 month ahead		12 month ahead	
	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative
Buying Conditions index	1.00	1.00	0.98	0.95*	0.99	0.93*	1.02	0.95***
ISM New Orders index	0.99	0.99	1.00	0.99	1.00	0.98	0.98	0.97*
Housing starts	1.02	0.96	1.03	0.97	1.03	0.94**	1.01	0.90**
ADS index	0.98	1.07	1.01	0.99	1.04	0.96	1.05	0.96*
Term spread	1.01	1.06	0.96	1.01	0.91**	0.92***	0.86***	0.87***
Default spread	1.01	1.16	1.16	1.12	1.39	1.06	1.44	1.04
3M T-bill rate	1.03*	0.99	1.02	1.01	1.03	1.01	1.03	1.00
Excess market return	0.98	1.18*	1.02	1.01	1.07*	1.02	1.16***	1.02
Realised volatility (RV)	1.02	1.25*	1.19	1.24	1.47	1.16	1.64	1.09
Principal component 1	0.99	0.98	1.03	0.99	1.09	0.97	1.08	0.97
Principal component 2	1.02	0.99	0.98	0.97	0.97	0.96**	0.96	0.94**
Principal component 3	1.03	1.08	0.98	1.01	1.00	0.96**	1.02	0.97**

Benchmark: GJR-GARCH(1,1). MAFE ratio: $\frac{MAFE_{GMX}}{MAFE_{GARCH}}$, where $MAFE_{GMX}$ is the mean absolute forecast error of the GARCH-MIDAS model driven by some economic data (X). A value below 1 means the GARCH-MIDAS model outperforms the GJR-GARCH(1,1) model. *, ** and *** indicate rejection of the null hypothesis of equal (unconditional) predictive ability at the 10%, 5%, and 1% level, respectively, according to the Giacomini and White (2006) test. Positive / negative growth months defined according to the sign of annualised monthly industrial production growth (manufacturing only, most recent value): 95 low growth and 163 high growth periods.

From Table 4 we can see that the GJR-GARCH(1,1) model is still difficult to beat at short horizons. Macroeconomic variables, the term spread and the second and third PCs do, however,

improve forecasts in negative growth periods over long horizons. This is in line with the results in Figure 6, where many of the macroeconomic variables improved forecast in particular during the latest recession. The GARCH-MIDAS model augmented by the term spread is also the best model in expansions over long horizons. The performance of GARCH-MIDAS models driven by other financial data remains weak. The main conclusions carry over to the MSFEs (Table A.5) and to using NBER recession periods instead (Table A.13), although the results are weaker in terms of statistical significance.

6.3.2 Volatility environment

I next divide the sample period based on the VIX index, and as a robustness check (see Appendix A.8 for the robustness checks and plots of the regimes) the St. Louis Fed Financial Stress Index (STLFSI)²⁵, to determine how the forecast accuracy of the GARCH-MIDAS models are impacted by the volatility environment. The results are presented in Table 5.

Table 5: Effect of volatility environment on forecasting performance: MAFE ratios

	1 month ahead		3 month ahead		6 month ahead		12 month ahead	
	Low vola	High vola	Low vola	High vola	Low vola	High vola	Low vola	High vola
Buying Conditions index	0.99	1.01	0.89**	0.98	0.83***	0.99	0.84**	1.01
ISM New Orders index	0.97	0.99	0.88**	1.02	0.84***	1.02	0.79***	1.02
Housing starts	0.97	0.99	0.91*	1.02	0.88**	1.00	0.84*	0.97
ADS index	0.99	1.03	0.96	1.01	0.94*	1.01	0.92*	1.02
Term spread	1.01	1.04	0.85*	1.02	0.71***	0.96	0.58***	0.94**
Default spread	0.99	1.11	0.97	1.18	1.00	1.25	1.14	1.22
3M T-bill rate	0.95***	1.02	0.84***	1.05***	0.80***	1.07***	0.76***	1.07***
Excess market return	0.97	1.11*	0.99	1.02	1.13**	1.02	1.20**	1.05*
Realised volatility (RV)	0.94	1.19**	0.83**	1.31**	0.89	1.38	1.22	1.33
Principal component 1	0.97	0.99	0.96	1.02	0.97	1.04	0.97	1.03
Principal component 2	0.94**	1.02	0.78***	1.02	0.68***	1.03	0.61***	1.03*
Principal component 3	0.98	1.07	0.83***	1.03*	0.76***	1.03	0.70***	1.06**

Benchmark: $GJR-GARCH(1,1)$ model. MAFE ratio: $\frac{MAFE_{GMX}}{MAFE_{GARCH}}$, where $MAFE_{GMX}$ stands for the mean absolute forecast error from the GARCH-MIDAS model driven by some macroeconomic or financial data (X). A value below 1 means the GARCH-MIDAS model outperforms the $GJR-GARCH(1,1)$ model. *, ** and *** indicate a rejection of the null hypothesis of equal (unconditional) predictive ability at the 10%, 5%, and 1% level, respectively, according to the Giacomini and White (2006) test. High / low volatility months are based on the median of the VIX index: 147 months of high volatility and 111 months of low volatility.

Many of the economic variables are useful for forecasting volatility in low volatility periods, while mainly failing to do so in high volatility periods even over long horizons. The low volatility periods take place right before the financial crisis in 2007-2008 and after roughly 2013 (see Figure A.23). We could already see from Figure 6 that macroeconomic data was useful for

²⁵The STLFSI consists of 18 series, including several interest rates, yield curves and the VIX index.

forecasting volatility roughly during these time periods. Thus the results in this section confirm that the differences in forecasting performance uncovered in Section 6.2 can at least partly be explained by changes in the volatility environment. Even on the 1M horizon there are now some statistically significant improvements over the baseline model in low volatility periods. Interestingly, the model driven by the 3M T-bill rate clearly improves forecasts in low volatility periods while leading to clearly worse forecasts in the high volatility periods. Especially the second and third principal components perform very well in low volatility environments. Thus clearly economic variables improve the accuracy of stock return volatility forecasts in low volatility periods. It seems intuitive that economic data is more important for forecasts during calm markets, while the GARCH model, which reacts more quickly to changes in the market environment, performs better in high volatility environments. The main results are robust to using mean squared forecast errors (Table A.7) and to using the financial stress indicator (Table A.14), although they are weaker in terms of statistical significance.²⁶ Clearly, if we could correctly anticipate being in a low volatility environment we might be able to improve volatility forecasts by including economic data in a GARCH model.

6.4 Conditional predictive ability

In the previous section I determined that relative forecasting performance depends on the business cycle and the volatility environment. This section takes the idea one step further and explores whether relative forecasting performance is predictable using information on the state of the economy or the volatility environment available at the forecast origin. This information could be exploited in forecast combination schemes or forecast model selection. I apply the conditional predictive ability test by Giacomini and White (2006), and statistically test whether relative forecasting performance is predictable using the (expected) state of the business cycle (Survey of Professional Forecaster data, real-time professional recession probabilities²⁷), or an indicator for financial market volatility (VIX index).²⁸ The interpretation of the test is such that if we find that the conditional test rejects while the unconditional test fails to reject (Table 3), then even though average performance is roughly equal, the relative performance could have been predicted using information on the economic or market environment at the forecast

²⁶These results do not get strong support from the MSFE ratios when dividing the sample based on the financial stress index (Table A.8), indicating low volatility rather than low stress is important.

²⁷Quarterly data transformed into monthly by keeping it fixed within each quarter.

²⁸As a robustness check/alternative measures I also use the NBER recession dates, industrial production growth, the STLFSI, and previous period forecast losses. See Appendix A.9 for details.

horizon. On the other hand, if the unconditional test rejects while the conditional test does not, then the conditional test could have low power or the unconditional test could be undersized (see Giacomini and White (2006)).

Table 6: Conditional test using SPF recession probabilities: MAFE ratios

	1M ahead	2M ahead	3M ahead	6M ahead	9M ahead	12M ahead
Buying Conditions index	1.00	0.98	0.96	0.96**	0.97**	0.98**
ISM New Orders index	0.99	1.00	0.99	0.99	0.98*	0.97*
Housing starts	0.99	0.99	1.00	0.97	0.95	0.94
ADS index	1.03	1.00	1.00	0.99	1.00	1.00
Term spread	1.03	1.03	0.99*	0.91***	0.88***	0.87***
Default spread	1.09	1.13	1.14	1.20	1.22	1.20
3M T-bill rate	1.01	1.02	1.02	1.02	1.02***	1.01
Excess market return	1.08	1.03	1.02	1.04*	1.06**	1.08**
Realised volatility (RV)	1.14	1.23*	1.22	1.29	1.32	1.31
First principal component	0.99	1.00	1.00	1.02	1.04	1.02*
Second principal component	1.01	0.99	0.98	0.96	0.96	0.95*
Third principal component	1.05	1.00	0.99	0.98*	0.98*	0.99

Benchmark: $GJR-GARCH(1,1)$. MAFE ratio: $\frac{MAFE_{GMX}}{MAFE_{GARCH}}$, where $MAFE_{GMX}$ stands for the mean absolute forecast error from the GARCH-MIDAS model driven by some macroeconomic or financial data (X). A value below 1 means the GARCH-MIDAS model outperforms the $GJR-GARCH(1,1)$ model. *, ** and *** indicate a rejection of the null hypothesis of equal (conditional) predictive ability at the 10%, 5%, and 1% level, respectively, according to the Giacomini and White (2006) test. Conditioning variable: Survey of Professional Forecasters recession probabilities. 1Q ahead for 1M to 3M ahead forecasts, 2Q ahead for 6M, 3Q ahead for 9M, and 4Q ahead for 12M ahead forecasts. Test function: $h_t = [1 \ v_t]$, where v_t is the conditioning information.

Table 7: Conditional test using the VIX index: MAFE ratios

	1M ahead	2M ahead	3M ahead	6M ahead	9M ahead	12M ahead
Buying Conditions index	1.00	0.98	0.96	0.96**	0.97**	0.98**
ISM New Orders index	0.99	1.00	0.99	0.99	0.98	0.97
Housing starts	0.99	0.99	1.00	0.97**	0.95***	0.94**
ADS index	1.03	1.00	1.00	0.99	1.00	1.00
Term spread	1.03	1.03	0.99	0.91***	0.88***	0.87***
Default spread	1.09	1.13	1.14	1.20	1.22	1.20
3M T-bill rate	1.01	1.02	1.02	1.02*	1.02***	1.01
Excess market return	1.08	1.03	1.02	1.04	1.06**	1.08**
Realised volatility (RV)	1.14	1.23*	1.22	1.29	1.32	1.31
First principal component	0.99	1.00	1.00	1.02	1.04	1.02
Second principal component	1.01	0.99	0.98	0.96	0.96***	0.95**
Third principal component	1.05	1.00	0.99	0.98	0.98**	0.99*

Benchmark: $GJR-GARCH(1,1)$. MAFE ratio: $\frac{MAFE_{GMX}}{MAFE_{GARCH}}$, where $MAFE_{GMX}$ stands for the mean absolute forecast error from the GARCH-MIDAS model driven by some macroeconomic or financial data (X). A value below 1 means the GARCH-MIDAS model outperforms the $GJR-GARCH(1,1)$ model. *, ** and *** indicate a rejection of the null hypothesis of equal (conditional) predictive ability at the 10%, 5%, and 1% level, respectively, according to the Giacomini and White (2006) test. Conditioning variable: level of VIX index. Test function: $h_t = [1 \ v_t]$, where v_t is the conditioning information.

Comparing the significance of the loss function differences in Table 6 to those in Table 3,

we can see that when using information on the (expected) business cycle, the forecast errors are predictable over long horizons for the GARCH-MIDAS model driven by the Buying Conditions index, and to a lesser degree when the ISM New Orders index is used. However, as is clear from Table 7, there is more predictability in forecast errors when using the volatility environment as the conditioning variable: the forecast errors are now predictable over long horizons when long-term volatility is driven by the Buying Conditions index, housing starts or the second or third PC.²⁹ Thus there is some evidence of predictability in forecast errors, especially when conditioning on the volatility environment. However, using a decision rule based on conditional predictive ability, as suggested by Giacomini and White (2006), does not lead to significant forecast improvements.³⁰

7 Forecast combination schemes

If the relative forecasting performance of models varies over time, forecast combination methods can be useful. The seminal paper by Bates and Granger (1969) already concluded that combination forecasts can outperform the individual forecasts, a conclusion widely confirmed in later literature.³¹ In practice, simple forecast combination methods, such as equal weights, often lead to more accurate forecasts than more complicated schemes (e.g., Clemen (1989))

This section explores combining forecasts produced by the GARCH-MIDAS models, using both simple and time-varying combination schemes. The simple combination schemes are the mean, the median and the trimmed mean of the GARCH-MIDAS forecasts, where the trimmed mean refers to removing the smallest and the largest forecasts each period and taking a mean of the remaining forecasts. Because the financial variables produced a clearly inferior forecast on all horizons over most time periods, I focus on combining the forecasts produced by the macroeconomic variables and the term spread.³² The time-varying alternatives either use time-varying weights (the discounted mean absolute (or square) prediction error (DMAPE/DMSPE) following Stock and Watson (2004)) or choose the forecast(s) to be used by ranking the forecast based on past performance, i.e., past forecast errors (similar to, e.g., Aiolfi and Timmermann (2006)).

²⁹These results are even slightly stronger when the STLFSI is used as the conditioning variable (Table A.17).

³⁰The simple decision rule used here adaptively selects either the GARCH-MIDAS based forecast or the baseline GJR-GARCH forecast, depending on whether equal conditional predictive ability can be rejected at the forecast origin or not, see Section 4 in Giacomini and White (2006) for details. The results are available upon request.

³¹See, for example, Clemen (1989), Chan et al. (1999) and Stock and Watson (1999).

³²Appendix A.10 discusses the forecast combination results using all the series. Overall the results are weaker when the generally inferior forecasts produced using financial data are included.

The DMSPE forecast combination scheme is used by, for example, Rapach et al. (2010) for equity premium prediction and Paye (2012) for stock market volatility forecasts in the predictive regression setting.

The combination forecasts are weighted averages of the N individual forecasts ($\hat{\sigma}_{i,t+1}^2$): $\hat{\sigma}_{c,t+1}^2 = \sum_{i=1}^N \omega_{i,t} \hat{\sigma}_{i,t+1}^2$, where the weights depend on the chosen combination method. For example, the simple mean combination puts $\omega_{i,t} = \frac{1}{N}$. The DMAPE weights depend on the historical performance of the models:

$$\omega_{i,t} = \frac{\phi_{i,t}^{-1}}{\sum_{j=1}^N \phi_{j,t}^{-1}}, \text{ where } \phi_{i,t} = \sum_{s=1}^{t-h} \eta^{t-h-s} |\sigma_{s+h}^2 - \hat{\sigma}_{i,s+h}^2|$$

and h is the forecasting horizon. $0 < \eta \leq 1$ is the discount factor: $\eta = 1$ is the basic case from Bates and Granger (1969) for uncorrelated individual forecasts. When $\eta < 1$ recent forecast accuracy is weighted more heavily. I use $\eta = 1$ and $\eta = 0.5$. Stock and Watson (2004) conclude that for macroeconomic forecasting more discounting ($\eta = 0.9$) usually performs at least no better than less discounting (e.g., $\eta = 1$).

Table 8: Combination forecasts: MAFE ratios

	1 month ahead	3 month ahead	6 month ahead	9 month ahead	12 month ahead
Mean	1.00	0.96**	0.95**	0.94**	0.94**
Median	0.99	0.97*	0.96*	0.96	0.94***
Trimmed mean	0.99	0.97*	0.95*	0.95*	0.94***
DMAPE, $\eta = 1$	1.00	0.96**	0.95**	0.94**	0.94***
DMAPE, $\eta = 0.5$	1.00	0.96**	0.94**	0.94**	0.93**
Previously best	1.05	0.98	0.91***	0.89***	0.87***
Mean (best three)	1.00	0.97	0.94***	0.94**	0.93***

*Benchmark: GJR-GARCH(1,1). MAFE ratio: $\frac{MAFE_{combo}}{MAFE_{GARCH}}$, where $MAFE_{combo}$ stands for the mean absolute forecast error from the combination forecast using the method stated in the first column. A value below 1 means the combination forecast outperforms the GJR-GARCH(1,1) model. *, ** and *** indicate a rejection of the null hypothesis of equal (unconditional) predictive ability at the 10%, 5%, and 1% level, respectively, according to the Giacomini and White (2006) test. The last four combination schemes are based on the forecasting performance over an expanding window of initial size 12 months. Note that due to initial calculations all forecast comparisons are for the period January 1998 - June 2017 (234 periods).*

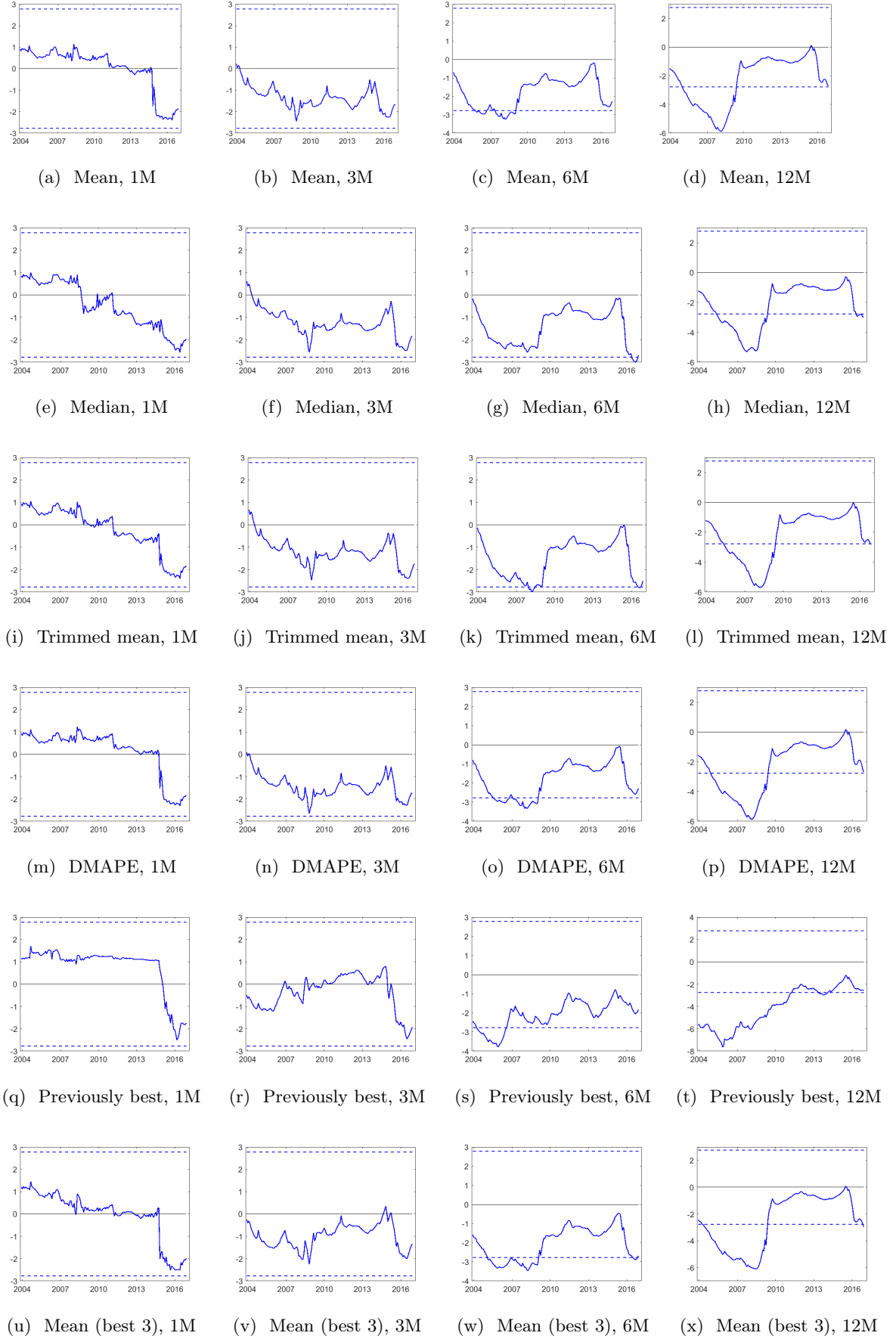


Figure 7: Fluctuation test applied to forecast combinations of the individual GARCH-MIDAS models. Dashed lines represent 10% confidence bands. Benchmark: GJR-GARCH(1,1) model. DMAPE with $\eta = 0.5$. Note that the year on the x-axis marks the end of the rolling window period, over which the test statistics is calculated. $m = 78$

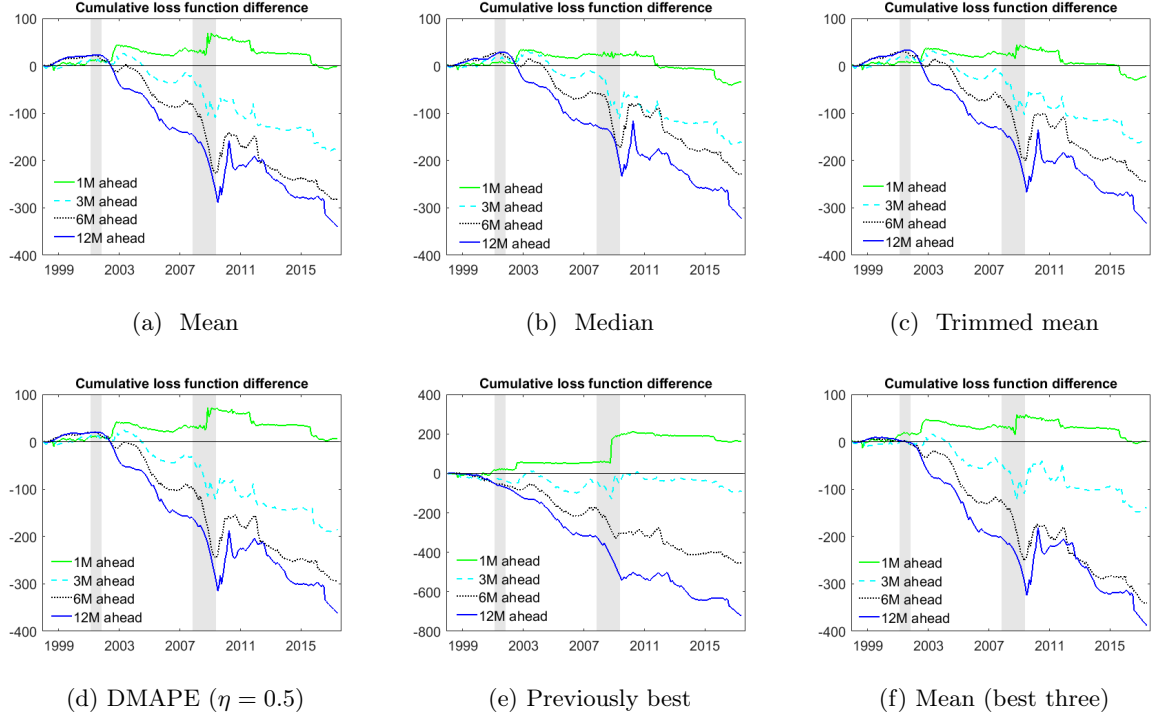


Figure 8: Cumulative sum of loss function differences (absolute errors) of forecast combinations of the individual GARCH-MIDAS models, compared to the GJR-GARCH(1,1) model ($|Loss_{combo} - Loss_{GARCH}|$). An upward sloping segment thus indicates the GJR-GARCH model outperforms the GARCH-MIDAS model. Grey areas mark NBER dated US recessions.

On the other hand, if there is clear persistence in forecasting performance and the differences between model accuracy are large, we can potentially improve on the simple mean by excluding the worst performing models in each period. Considering the results in the previous section it is clear that there were some models which produced inferior forecasts for a prolonged period of time, and preselecting the included forecasting models based on past performance can thus be beneficial. I rank the forecasting models in each period and for each horizon based on average past performance over an expanding window with initial size of 12 months. In each out-of-sample period I then pick the forecast of the model that has had the best average forecasting performance up until the forecast origin ('Previously best'), as well as take the mean of the forecasts of the best-performing three models ('Mean (best three)').

Table 8 gives the mean absolute forecast error ratios of the combination forecasts. Over the 1 month horizon performance is similar to the benchmark. Over longer horizons the forecast combinations tend to significantly outperform the benchmark, contrary to most of the individual forecasts. The Fluctuation test, which tests whether the forecasting performance is time-varying, reveals that the test statistics are, especially on horizons longer than 1 month, predominantly

negative, and the GJR-GARCH(1,1) never significantly outperforms any of the combination forecasts (see Figure 7 for the results). Thus the combination forecasts outperform most of the individual forecasts in more consistently outperforming the benchmark model. During the first half of the sample the differences in forecasting performance are often statistically significant on the 12 month horizon, but also to a lesser degree on the 6 month horizon, in favour of the combination methods. We can see that most of the combination schemes produce qualitatively similar forecasts, implying that it does not greatly matter whether a simple or a time-varying combination scheme is chosen.³³ The exception is the combination scheme using only the forecast of the best performing model, which on longer horizons largely replicates the performance of the term spread driven GARCH-MIDAS model.

In addition, many of the combination schemes have recently performed well against the benchmark model, although the difference is slightly statistically significant only for the median and the 'Mean (best three)' on the 6 and 12 month horizons. The performance of most of the combination methods seems to have deteriorated around the financial crisis, which is reflected over the whole 6.5 year period for the Fluctuation test. Comparing the forecast combinations to the principal component driven models (see Appendix A.7 for results for the PC driven models) reveals that forecast combinations perform better than the models using information from a large amount of economic data.

To shed further light on the performance of the combination forecasts I plot the cumulative sum of the loss function differences in Figure 8. The period of weak performance for most of the combination schemes, evident in the Fluctuation test statistics, is confirmed to stem mainly from weak performance immediately after the latest recession and financial crisis. Thus, forecast combinations seem useful for forecasting volatility in many periods and provide forecasts that are consistently at least slightly better than the benchmark model, for horizons longer than one month. The mean squared forecast errors, presented in the appendix in Table A.4, suggest qualitatively similar conclusions, but the results are mostly not statistically significant.

8 Conclusion

This paper evaluates the time-variation in the relative forecasting performance of models for stock return volatility, with focus on using macroeconomic and financial data to enhance long-

³³The number of models being combined is fairly modest (5), and a larger amount of individual models could reveal larger differences between the different combination schemes.

horizon volatility forecasts. The paper addresses three related questions. First, how does the predictive ability of different economic variables vary over time? Second, does the economic environment affect the forecasting performance of different variables? Third, can forecast accuracy be improved by combining individual GARCH-MIDAS forecasts?

When forecasting over long horizons there are clear shifts in forecasting performance over time implying that (time-varying) forecast combination methods could be useful. Macroeconomic variables improve predictions especially in low volatility periods but also in periods of weak economic growth, while financial data driven GARCH-MIDAS models struggle to outperform the benchmark GJR-GARCH model. However, although some forecast errors are predictable conditioning on especially the volatility environment, it is difficult to achieve significant improvements in forecast accuracy in real-time forecasting. It is clear that no single forecasting model or combination scheme performs well on all horizons and in all time periods. When forecasting 12 months ahead the best forecasting model is the term spread driven GARCH-MIDAS model, while when forecasting 3 or 6 months ahead you would most likely choose a forecast combination method. Over the 1 month horizon there is some evidence that a GARCH-MIDAS model, or a combination method, currently performs best, although there are no statistically significant differences and the GJR-GARCH model performed well especially in the first half of the sample period. As the GJR-GARCH model is rarely significantly better than the GARCH-MIDAS models and never significantly outperforms the combination forecasts, it is useful to augment the model with economic data for long horizon forecasts. This paper only briefly considered using conditional predictive ability to improve forecast accuracy. An interesting question for future research is establishing whether a forecast selection method, which consistently and significantly outperforms the GJR-GARCH model, exists.

Bibliography

- Aiolfi, Marco, and Allan Timmermann (2006) ‘Persistence in forecasting performance and conditional combination strategies.’ *Journal of Econometrics* 135, 31–53
- Andersen, Torben G., and Tim Bollerslev (1998) ‘Answering the skeptics: Yes, standard volatility models do provide accurate forecasts.’ *International Economic Review* 39(4), 885–905.
- Andreou, Elena, Eric Ghysels, and Andros Kourtellis (2010) ‘Regression models with mixed sampling frequencies.’ *Journal of Econometrics* 158, 246–261.
- Asgharian, Hossein, Ai Jun Hou, and Farrukh Javed (2013) ‘The importance of the macroeconomic variables in forecasting stock return variance: a GARCH-MIDAS approach.’ *Journal of Forecasting* 32, 600–612.
- Asgharian, Hossein, Charlotte Christiansen, and Ai Jun Hou (2015) ‘Effects of macroeconomic uncertainty on the stock and bond markets.’ *Finance Research Letters* 13, 10–16.
- Awartani, Basel M. A., and Valentina Corradi (2005) ‘Predicting the volatility of the S&P-500 stock index via GARCH models: the role of asymmetries.’ *International Journal of Forecasting* 21, 167–183.
- Baker, Malcolm, and Jeffrey Wurgler (2006) ‘Investor sentiment and the cross-section of stock returns.’ *The Journal of Finance* 61(4), 1645–1680
- Bates, John M, and Clive WJ Granger (1969) ‘The combination of forecasts.’ *Or* pp. 451–468
- Bollerslev, Tim (1986) ‘Generalized autoregressive conditional heteroskedasticity.’ *Journal of Econometrics* 31, 307–327.
- Cakmakli, Cem, and Dick van Dijk (2010) ‘Getting the most out of macroeconomic information for predicting stock returns and volatility.’ *Tinbergen Institute Discussion paper, No. 10-115/4*
- Campbell, John Y. (1991) ‘A variance decomposition for stock returns.’ *Economic Journal* 101, 157–179.
- Campbell, John Y., and Robert J. Shiller (1988) ‘The dividend-price ratio and expectations of future dividends and discount factors.’ *Review of Financial Studies* 1, 195–228.

- Chan, Yeung Lewis, James H. Stock, and Mark W. Watson (1999) ‘A dynamic factor model framework for forecast combination.’ *Spanish Economic Review* 1, 91–121.
- Chauvet, Marcelle, and Simon Potter (2013) ‘Forecasting output.’ In ‘Handbook of Economic Forecasting,’ vol. 2 (Elsevier) pp. 141–194
- Christiansen, Charlotte, Maik Schmeling, and Andreas Schrimpf (2012) ‘A comprehensive look at financial volatility prediction by economic variables.’ *Journal of Applied Econometrics* 27(6), 956–977.
- Christoffersen, Peter F., and Francis X. Diebold (2000) ‘How relevant is volatility forecasting for financial risk management?’ *The Review of Economics and Statistics* 82(1), 12–22.
- Clark, T., and K.D. West (2006) ‘Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis.’ *Journal of Econometrics* 135, 155–186.
- Clark, Todd E., and Michael W. McCracken (2001) ‘Tests of equal forecast accuracy and encompassing for nested models.’ *Journal of Econometrics* 105(1), 85–110.
- Clemen, Robert T. (1989) ‘Combining forecasts: A review and annotated bibliography.’ *International Journal of Forecasting* 5, 559–581.
- Conrad, Christian, and Karen Loch (2014) ‘Anticipating long-term stock market volatility.’ *Journal of Applied Econometrics* 30(7), 1090–1114.
- Davis, Nicole, and Ali M. Kutan (2003) ‘Inflation and output as predictors of stock returns and volatility: international evidence.’ *Applied Financial Economics* 13(9), 693–700.
- Diebold, Francis X., and Kamil Yilmaz (2008) ‘Macroeconomic volatility and stock market volatility, world-wide.’ *NBER Working Paper*
- Diebold, Francis X., and Robert S. Mariano (1995) ‘Comparing predictive accuracy.’ *Journal of Business and Economic Statistics* 13, 253–263.
- Engle, Robert F. (1982) ‘Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation.’ *Econometrica* 50(4), 987–1008.
- Engle, Robert F., and Gary Lee (1999) *A permanent and transitory component model of stock return volatility*, R. Engle and H. White (ed.) *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W. J. Granger* (Oxford, UK: Oxford University Press)

- Engle, Robert F., and Jose G. Rangel (2008) ‘The SPLINE-GARCH model for low frequency volatility and its global macroeconomic causes.’ *Review of Financial Studies* 21, 1187–1222.
- Engle, Robert F., Eric Ghysels, and Bumjean Sohn (2013) ‘Stock market volatility and macroeconomic fundamentals.’ *The Review of Economics and Statistics* 95(3), 776–797.
- Errunza, Vihang, and Ked Hogan (1998) ‘Macroeconomic determinants of European stock market volatility.’ *European Financial Management* 4(3), 361–377.
- Ghysels, Eric, Antonio Rubia, and Rossen Valkanov (2009) ‘Multi-period forecasts of volatility: Direct, iterated and mixed-data approaches.’ *working paper*
- Ghysels, Eric, Arthur Sinko, and Rossen Valkanov (2007) ‘MIDAS regressions: Further results and new directions.’ *Econometric Reviews* 26(1), 53–90.
- Ghysels, Eric, Pedro Santa-Clara, and Rossen Valkanov (2004) ‘The MIDAS touch: Mixed data sampling regression models.’ *Finance*.
- (2005) ‘There is a risk-return tradeoff after all.’ *Journal of Financial Economics* (76), 509–548.
- (2006) ‘Predicting volatility: getting the most out of return data sampled at different frequencies.’ *Journal of Econometrics* 131(1-2), 59–95.
- Giacomini, Raffaella, and Barbara Rossi (2010) ‘Forecast comparisons in unstable environments.’ *Journal of Applied Econometrics* 25, 595–620.
- Giacomini, Raffaella, and Halbert White (2006) ‘Tests of conditional predictive ability.’ *Econometrica* 74, 1545–1578.
- Glosten, Lawrence R., Ravi Jagannathan, and David E. Runkle (1993) ‘On the relation between the expected value and the volatility of the nominal excess return on stocks.’ *The Journal of Finance* 48(5), 1779–1801.
- Hamilton, James D., and Gang Lin (1996) ‘Stock market volatility and the business cycle.’ *Journal of Applied Econometrics* 11(5), 573–593.
- Hansen, Peter R., and Asger Lunde (2005) ‘A forecast comparison of volatility models: Does anything beat a GARCH(1,1)?’ *Journal of Applied Econometrics* 20, 873–889.

- Lindblad, Annika (2017) ‘Sentiment indicators and macroeconomic data as drivers for low-frequency stock market volatility.’ *HECER Discussion paper*
- McCracken, Michael W. (2000) ‘Robust out-of-sample inference.’ *Journal of Econometrics* 99(2), 195–223.
- McCracken, Michael W, and Serena Ng (2016) ‘Fred-md: A monthly database for macroeconomic research.’ *Journal of Business & Economic Statistics* 34(4), 574–589
- Patton, Andrew J. (2011) ‘Volatility forecast comparison using imperfect volatility proxies.’ *Journal of Econometrics* 160, 246–256.
- Paye, Bradley S. (2012) ‘Deja vol: Predictive regressions for aggregate stock market volatility using macroeconomic variables.’ *Journal of Financial Economics* 106, 527–546.
- Pierdzioch, Christian, Jorg Dopke, and Daniel Hartmann (2008) ‘Forecasting stock market volatility with macroeconomic variables in real time.’ *Journal of Economics and Business* 60, 256–276.
- Poon, Ser-Huang, and Clive W.J. Granger (2003) ‘Forecasting volatility in financial markets: A review.’ *Journal of Economic Literature* 41, 478–539.
- Rapach, David E, Jack K Strauss, and Guofu Zhou (2010) ‘Out-of-sample equity premium prediction: Combination forecasts and links to the real economy.’ *The Review of Financial Studies* 23(2), 821–862
- Schwert, William G. (1989) ‘Business cycles, financial crises, and stock volatility.’ *Carnegie-Rochester Conference Series on Public Policy* 31, 83–126.
- Stock, James H., and Mark W. Watson (1999) ‘Forecasting inflation.’ *Journal of Monetary Economics* 44, 293–375.
- (2003) ‘Forecasting output and inflation: the role of asset prices.’ *Journal of Economic Literature* 41, 788–829.
- Stock, James H, and Mark W Watson (2004) ‘Combination forecasts of output growth in a seven-country data set.’ *Journal of Forecasting* 23(6), 405–430
- Wang, Fangfang, and Eric Ghysels (2015) ‘Econometric analysis of volatility component models.’ *Econometric Theory* 31, 362–393.

West, Kenneth D. (1996) ‘Asymptotic inference about predictive ability.’ *Econometrica* 64, 1067–1084.

A Appendix

A.1 Data sources

- CRSP index: Kenneth French's Data Library
- ISM New Orders index: FRED database and the Institute for Supply Management (<https://www.instituteforsupplymanagement.org/>)
- Buying Conditions index: the University of Michigan consumer confidence report (<https://data.sca.isr.umich.edu/>)
- Housing starts: Philadelphia Fed real time center
- ADS index: Philadelphia Fed real time center
- Interest rates (including term spread): FRED database
- Default spread: St. Louis Fed, FRED database
- Excess market returns: Kenneth French's Data Library
- VIX index: St. Louis Fed, FRED database
- St. Louis Fed Financial Stress Index (STLFSI): St. Louis Fed, FRED database
- Survey of Professional Forecaster data, real-time professional recession probabilities: Philadelphia Fed real time center
- NBER recession dates: NBER (<http://www.nber.org/cycles.html>)

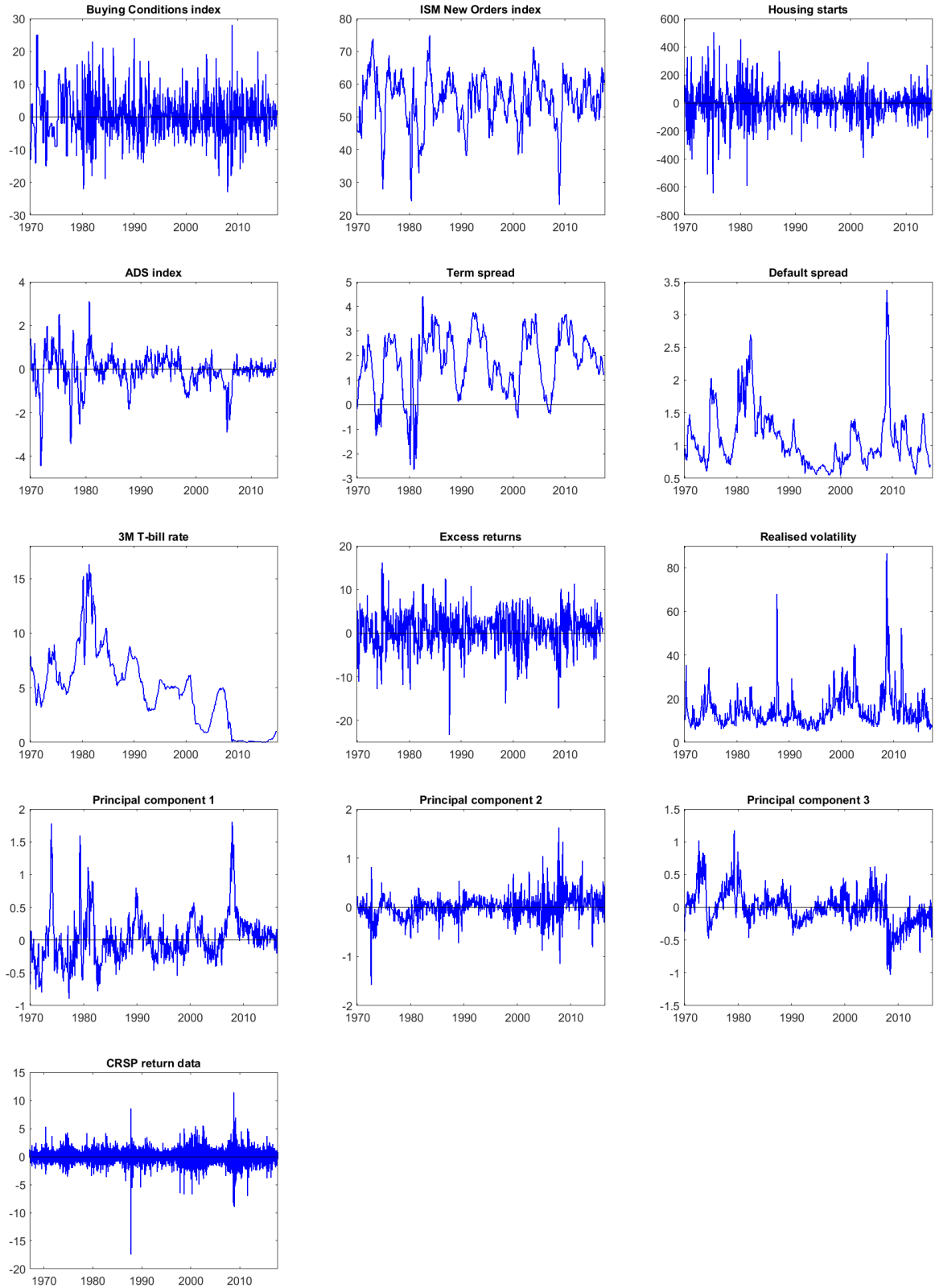


Figure A.1: Plots of explanatory data and return data used in the out-of-sample analysis.

A.2 PCA, marginal R^2 s for first four factors, full sample

This appendix presents the ten highest marginal R^2 s for the first four factors, extracted from the FRED-MD dataset. See McCracken and Ng (2016) for details on the data and the methodology to calculate the PCs.

The numbers in the parentheses denote the marginal R^2 for each factor, i.e., how much each factor explains of the overall variation in the data.

Table A.1: Ten highest marginal R^2 s for the first four factors

PC 1 (0.1472)		PC 3 (0.0685)	
Employment: Goods-Prod. Industries (USGOOD)	0.7106	Moody's Aaa Corporate Bond - Fed Funds (AAAFFM)	0.4487
Total nonfarm employment (PAYEMS)	0.7101	10Y Treasury C - Fed Funds (T10YFFM)	0.4438
IP: Manufacturing (SIC) (IPMANSICS)	0.6888	Moody's Baa Corporate Bond - Fed Funds (BAAFFM)	0.4333
IP Index (INDPRO)	0.6552	5Y Treasury C - Fed Funds (T5YFFM)	0.3956
Employment: Manufacturing (MANEMP)	0.6512	3M Treasury C - Fed Funds (TB3SMFFM)	0.3290
IP: Final Products and Nonindustrial Supplies (IPFPNSS)	0.6116	6M Treasury C - Fed Funds (TB6SMFFM)	0.3118
Employment: Durable goods (DMANEMP)	0.6001	1Y Treasury C - Fed Funds (T1YFFM)	0.2648
Capacity Utilization (manufacturing) (CUMFNS)	0.5927	CPI: Commodities (CUSR0000SAC)	0.2467
IP: Final Products (Market Group) (IPFINAL)	0.5137	Pers. Cons. Exp: Nondur. goods (DNDGRG3M086SBEA)	0.2437
IP: Durable Materials (SRVPRD)	0.4803	CPI (excl. shelter) (CUUR0000SA0L2)	0.2383
PC 2 (0.0708)		PC 4 (0.0558)	
CPI: Commodities (CUSR0000SAC)	0.5680	1Y Treasury Rate (GS1)	0.5073
Personal Cons. Exp. (Nondur.) (DNDGR3M086SBEA)	0.5573	5Y Treasury Rate (GS5)	0.4922
CPI (excl. shelter) (CUUR0000SA0L2)	0.5441	Moody's Seasoned Aaa Corporate Bond Yield (AAA)	0.4830
CPI: All Items (CPIAUCSL)	0.5321	6M Treasury Bill (TB6MS)	0.4707
CPI (excl. medical care) (CUSR0000SA0L5)	0.5016	10Y Treasury Rate (GS10)	0.4537
Personal Cons. Expenditure: Chain index (PCEPI)	0.4762	Moody's Seasoned Baa Corporate Bond Yield (BAA)	0.4374
CPI: Transportation (CPITRNSL)	0.4702	3M Treasury Bill: (TB3MS)	0.3749
CPI (excl. food) (CPIULFSL)	0.4299	3M AA Financial Commercial Paper Rate (CP3Mx)	0.3749
PPI: Finished Consumer Goods (PPIFCG)	0.3121	New Orders for Consumer Goods (ACOGNO)	0.2009
PPI: Finished goods (PPIFGS)	0.3595	S&P's Comp. Common Stock: Div. Yield (S&P div yield)	0.1864

Sample period: M12 1959 - M5 2017. Data set is the FRED-MD dataset, vintage June 2017 by McCracken and Ng (2016).

A.3 Rolling window principal components analysis (PCA)

This appendix presents rolling window results for the principal components analysis, detailing which series are most often chosen into the first three principal components. See McCracken and Ng (2016) for details on the series, and to link the series number to the name of the series.

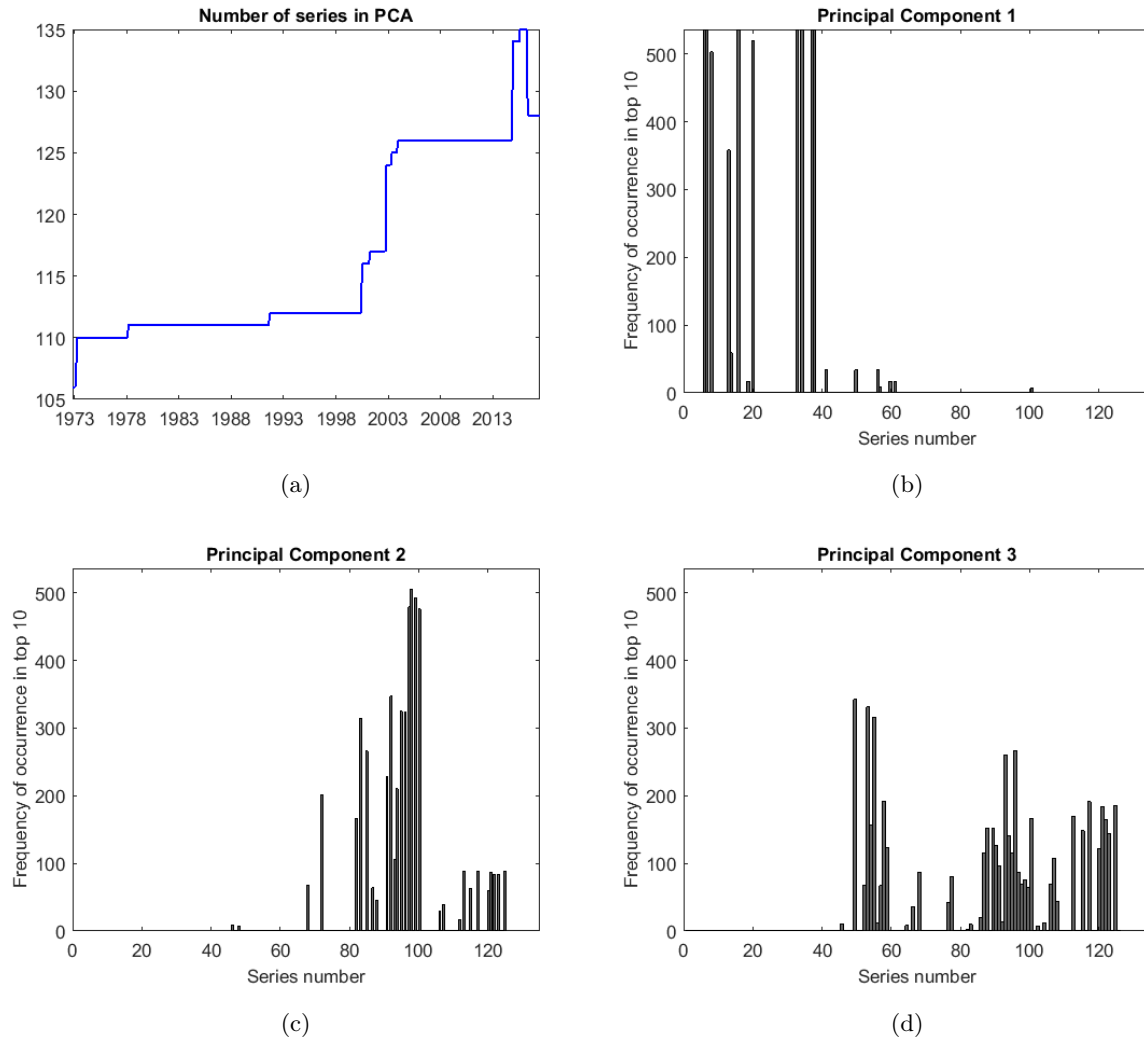


Figure A.2: Time-varying composition of the first three PCs. Panel (a) shows how the number of series in the data set varies over time.

A.4 Robustness check: Choice between restricted and unrestricted model

This appendix explores the implications of estimating one or two weights in the MIDAS polynomial. This is especially crucial for the term spread, excess market returns, PC2 and PC3, for which the choice of the optimal weighting schemes varies over time.

Figure A.3 shows the time variation in p-values from the likelihood ratio test between a model with one or two weights for each GARCH-MIDAS model. It shows that for the model

driven by the term spread or the second PC two weights have been preferred lately, while the opposite is true for the excess market return and the third PC.

Figure A.4 plots the estimates for θ , which are mostly similar for the two different choices of weighting schemes, except when the model with two weights is clearly superior.

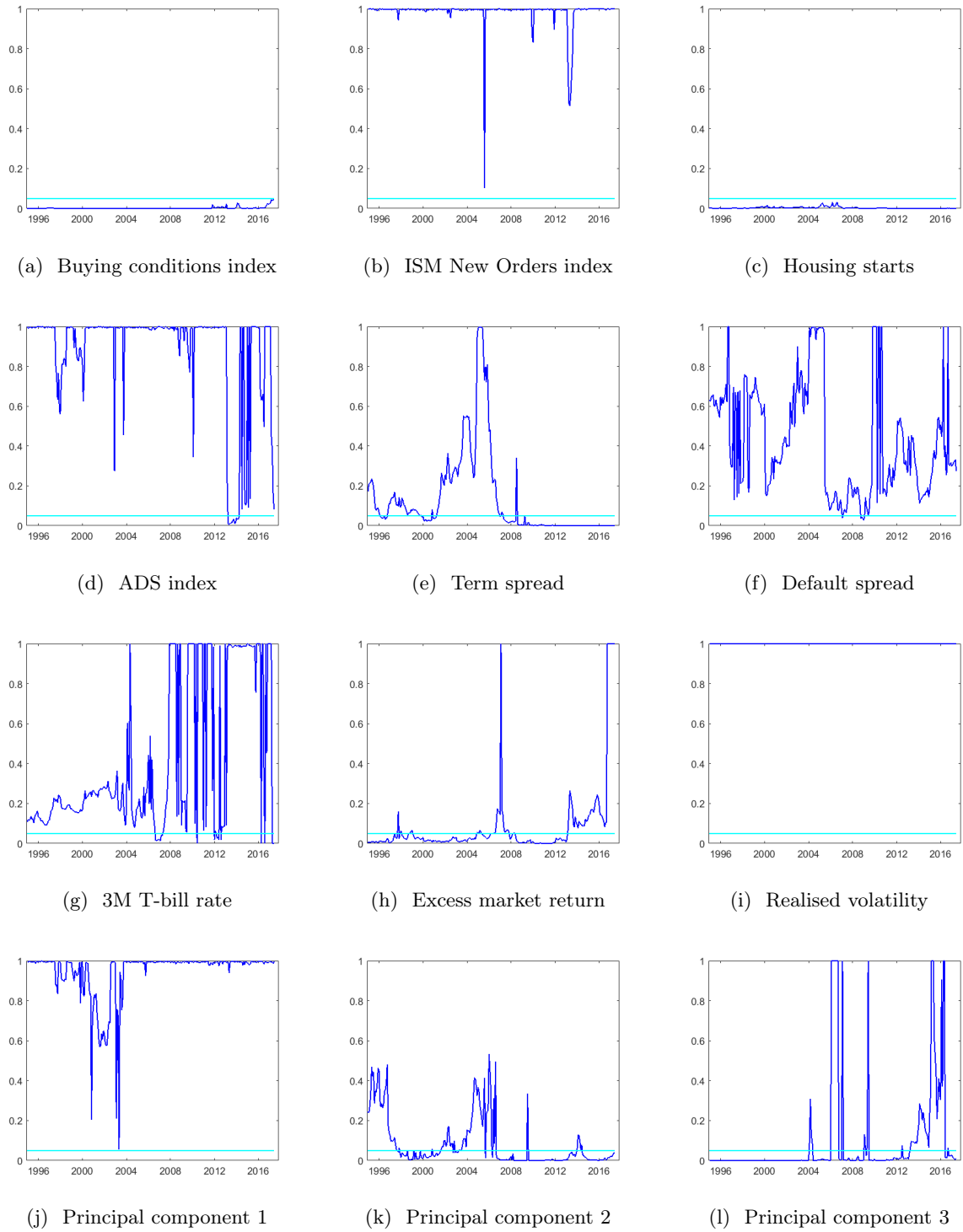
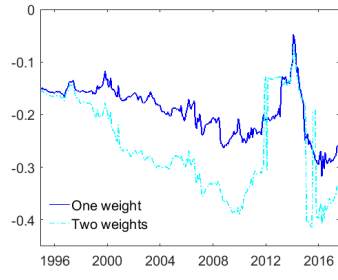
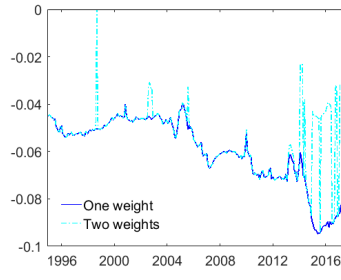


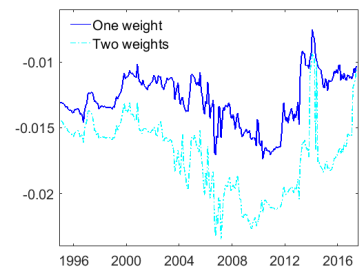
Figure A.3: p-values from the likelihood ratio test between the restricted and unrestricted models. Horizontal light blue line indicates the 5% significance level.



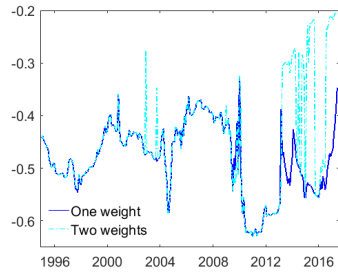
(a) Buying conditions index



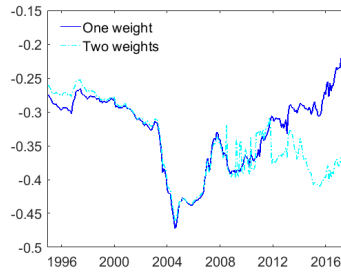
(b) ISM New Orders index



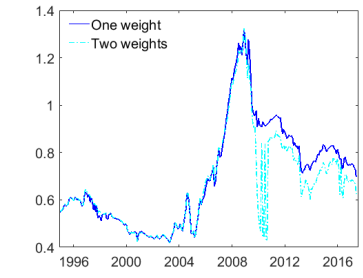
(c) Housing starts



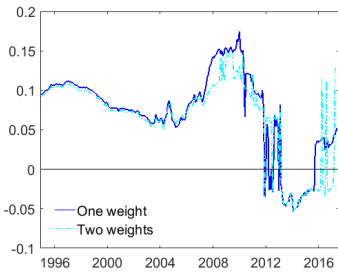
(d) ADS index



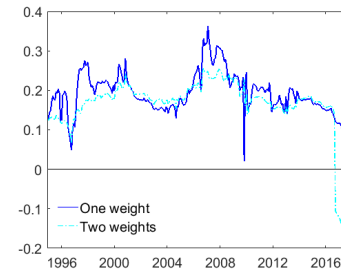
(e) Term spread



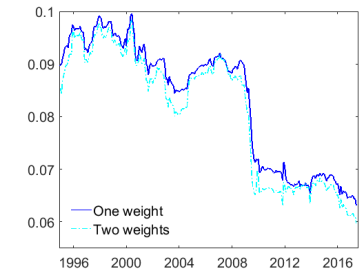
(f) Default spread



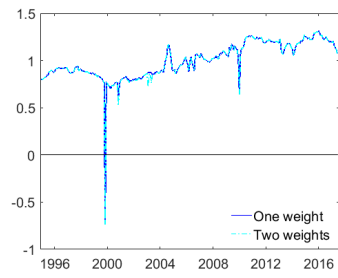
(g) 3M T-bill rate



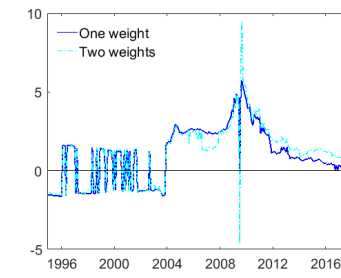
(h) Excess market return



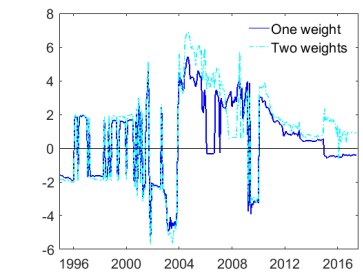
(i) Realised volatility



(j) Principal component 1



(k) Principal component 2



(l) Principal component 3

Figure A.4: Rolling window estimates of θ from the MIDAS polynomial with one or two weights.

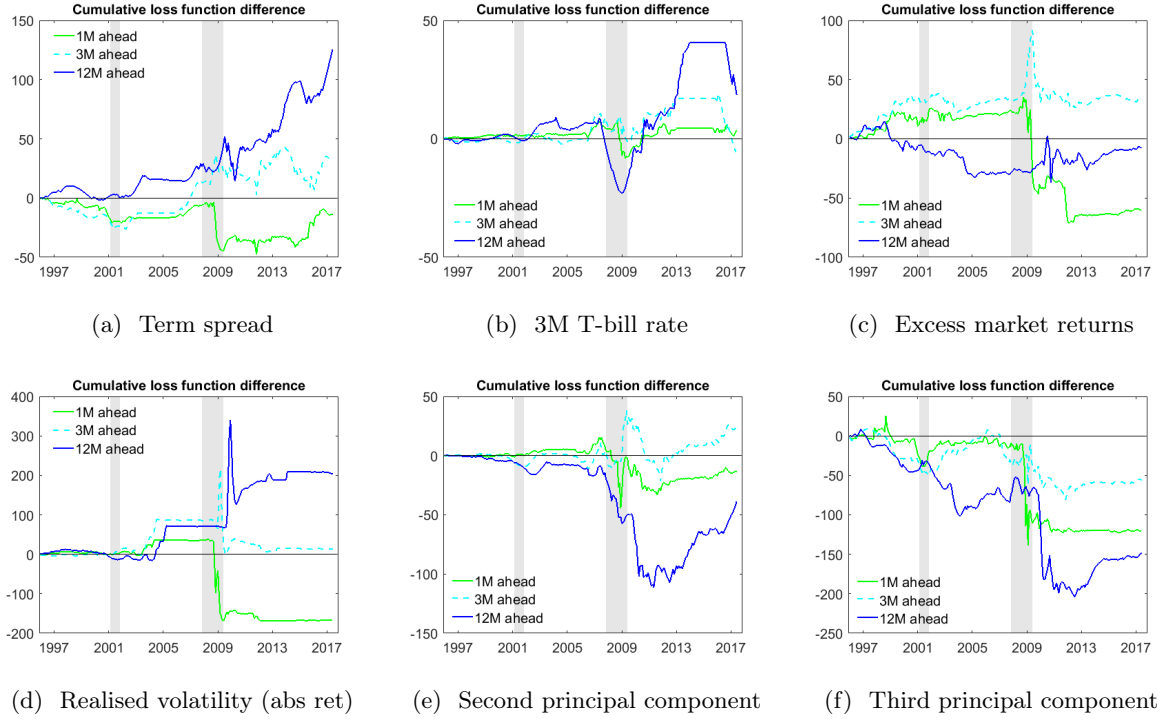


Figure A.5: Cumulative forecast loss differences between models estimated using one weight and two weights. An upward sloping line indicates the model estimated using two weights is superior.

Table A.2: MAFE ratios between models with one or two weight parameters

	1 month ahead	3 month ahead	6 month ahead	12 month ahead
Term spread	1.00	1.01	1.03	1.03
3M T-bill rate	1.00	1.00	1.00	1.00
Excess market return	0.98	1.01	0.99	1.00
Realised volatility (RV)	0.96	1.00	1.02	1.03
Second principal component	1.00	1.01	1.00	0.99
Third principal component	0.97	0.99	0.99	0.97

Benchmark: $GJR-GARCH(1,1)$. The MAFE ratios take the form: $\frac{MAFE_{GMX1w}}{MAFE_{GMX2w}}$, where $MAFE_{GMX1w}$ stands for the mean absolute forecast error from the GARCH-MIDAS model driven by some macroeconomic or financial data (X) with one estimated weight parameter in the MIDAS polynomial. A value below 1 means the GARCH-MIDAS model with one estimated weight parameter outperforms the GARCH-MIDAS model with two estimated weights parameters. $RV_t = \sum_{i=1}^{N_t} |r_{i,t}|$.

Figure A.5 and Table A.2 consider how the forecast errors change in the models estimated with one or two weight parameters (only the models for which the choice is ambiguous are considered). Over the full sample the differences in forecasting performance seem small, as the MAFE ratios are close to one. Over time the differences in forecasting performance vary, but the differences remain modest.

A.5 Mean squared forecast errors

This appendix presents results for squared forecast errors, as a robustness check to the absolute forecast errors presented in the main text. Despite the MSFE ratios being often relatively far from one, the predictive ability test by Giacomini and White (2006) and the Fluctuation test by Giacomini and Rossi (2010) mostly fail to reject the null hypothesis of equal predictive ability. The ranking of the models is, however, largely similar to the MAFE ratios. In the Fluctuation test (Figures A.6 and A.7) the recent performance of the GARCH-MIDAS models has been less convincing when looking at MSFEs than MAFEs.

Table A.3: Full sample results: Mean squared forecast error ratios

	1M ahead	2M ahead	3M ahead	6M ahead	9M ahead	12M ahead
Buying Conditions index	0.98	0.94	0.91	0.92	0.95	0.98
ISM New Orders index	1.00	1.00	1.00	1.00	0.99	0.99
Housing starts	0.98	0.97	0.96	0.93	0.90	0.90
ADS index	1.12	0.92	0.96	0.96	0.97	0.98
Term spread	1.22	1.07	1.00	0.94	0.91	0.91
Default spread	1.23	1.12	1.12	1.21	1.25	1.26
3M T-bill rate	1.04*	1.03**	1.04***	1.04***	1.03***	1.03**
Excess market return	1.36*	1.04	1.01	1.02	1.04	1.04
Realised volatility (RV)	1.24	1.32**	1.28	1.43	1.48	1.49
First principal component	0.88	0.95	0.96	0.97	0.98	0.98
Second principal component	1.04	1.00	1.00	0.98	0.97	0.97
Third principal component	1.21	1.03	1.00	0.98	1.01	1.00

Benchmark: $GJR-GARCH(1,1)$. The MSFE ratios take the form: $\frac{MSFE_{GMX}}{MSFE_{GARCH}}$, where $MSFE_{GMX}$ stands for the mean squared forecast error from the GARCH-MIDAS model driven by some macroeconomic or financial data (X). A value below 1 means the GARCH-MIDAS model outperforms the $GJR-GARCH(1,1)$ model. *, ** and *** indicate a rejection of the null hypothesis of equal (unconditional) predictive ability at the 10%, 5%, and 1% level, respectively, according to the Giacomini and White (2006) test. $RV_t = \sum_{i=1}^{N_t} |r_{i,t}|$.

Table A.4: Combination forecasts: MSFE ratios

	1 month ahead	3 month ahead	6 month ahead	9 month ahead	12 month ahead
Mean	1.04	0.96	0.94	0.94	0.95
Median	0.99	0.96	0.94	0.95	0.96*
Trimmed mean	1.01	0.96	0.94	0.94	0.95
DMSPE, $\eta = 1$	1.04	0.96	0.94	0.94	0.95
DMSPE, $\eta = 0.5$	1.04	0.95	0.94	0.93	0.94
Previously best	1.13	0.91	0.94	0.91	0.91*
Mean (best three)	1.02	0.94	0.94	0.95	0.95

Benchmark: $GJR-GARCH(1,1)$. MSFE ratio: $\frac{MSFE_{combo}}{MSFE_{GARCH}}$, where $MSFE_{combo}$ stands for the mean squared forecast error from the combination forecast using the method stated in the first column. A value below 1 means the combination forecast outperforms the $GJR-GARCH(1,1)$ model. *, ** and *** indicate a rejection of the null hypothesis of equal (unconditional) predictive ability at the 10%, 5%, and 1% level, respectively, according to the Giacomini and White (2006) test. The last four combination schemes are based on the forecasting performance over an expanding window of initial size 12 months. Note that due to initial calculations all forecast comparisons are for the period January 1998 - June 2017 (234 periods).

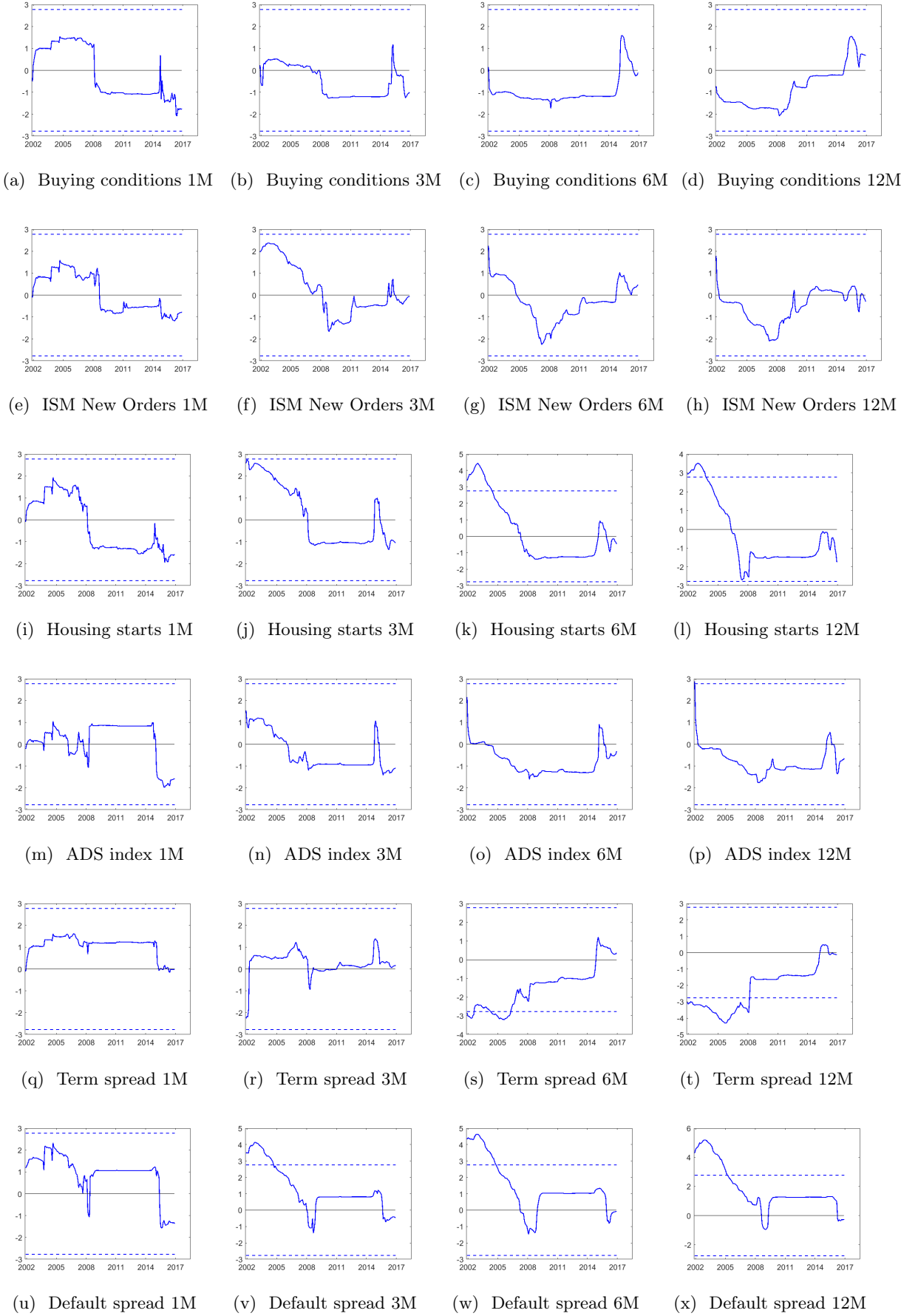


Figure A.6: Fluctuation test results for selected loss function differences. Squared forecast errors. Dashed lines represent 10% confidence bands. Note that the year on the x-axis marks the end of the rolling window period, over which the test statistics is calculated. Benchmark: GJR-GARCH(1,1) model. $l = 5$, $m = 78$

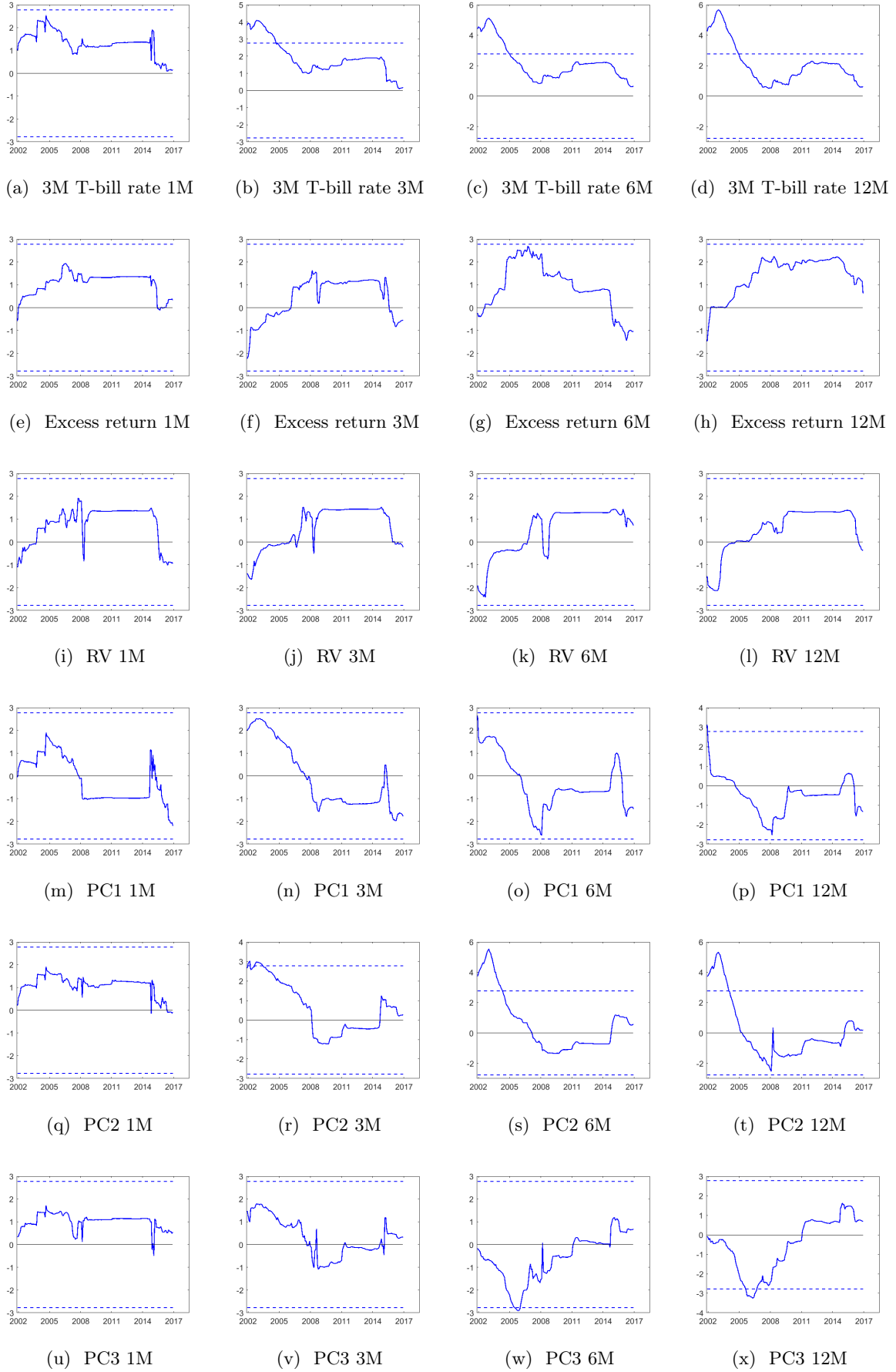


Figure A.7: Fluctuation test results for selected loss function differences. Squared forecast errors. Dashed lines represent 10% confidence bands. Note that the year on the x-axis marks the end of the rolling window period, over which the test statistics is calculated. Benchmark: GJR-GARCH(1,1) model. $l = 5$, $m = 78$

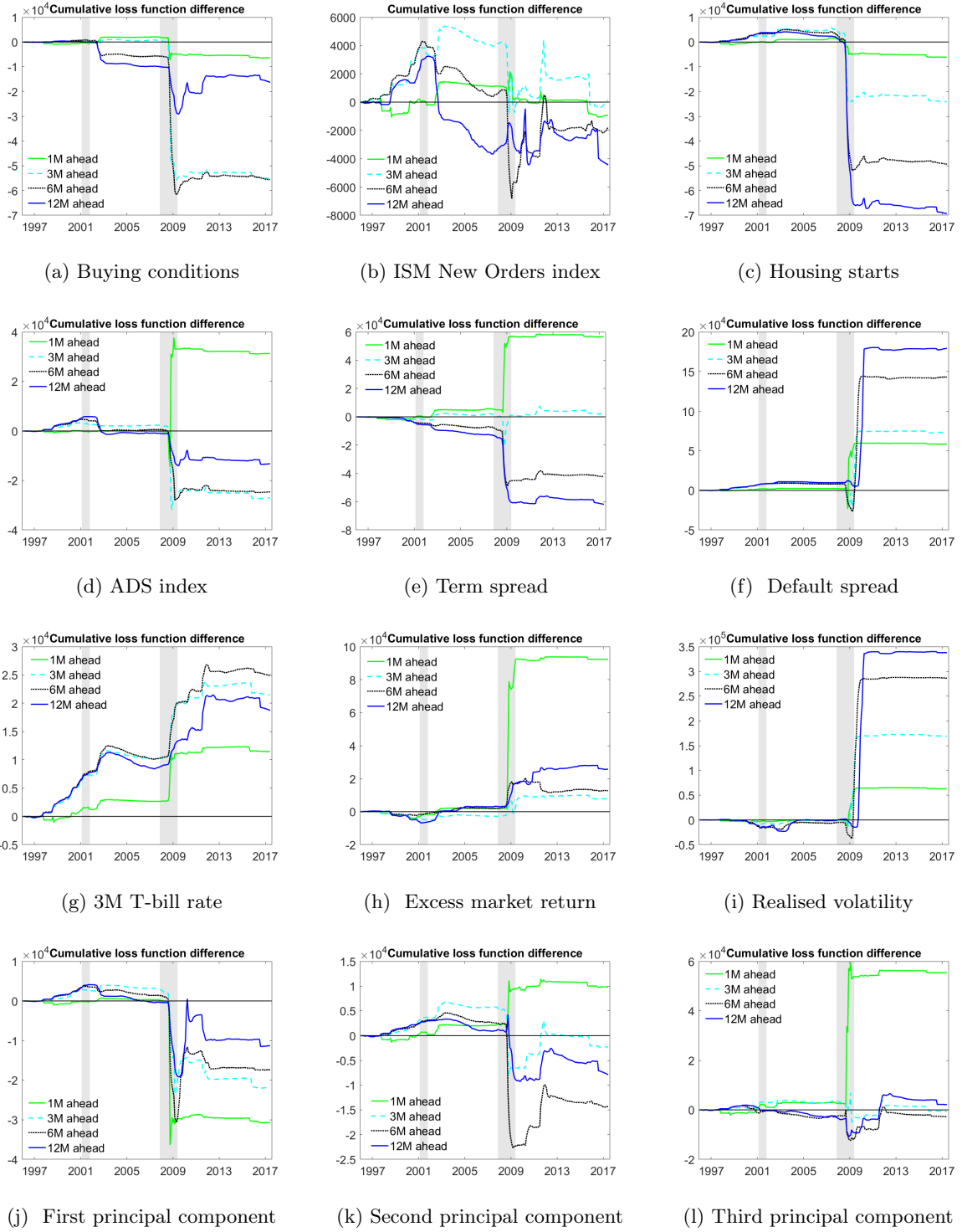


Figure A.8: Cumulative sum of loss function differences (squared errors) of forecast combinations of the individual GARCH-MIDAS models, compared to the GJR-GARCH(1,1) model ($(Loss_{combo} - Loss_{GARCH})^2$). An upward sloping segment thus indicates the GJR-GARCH model outperforms the GARCH-MIDAS model. Grey areas mark NBER dated US recessions.

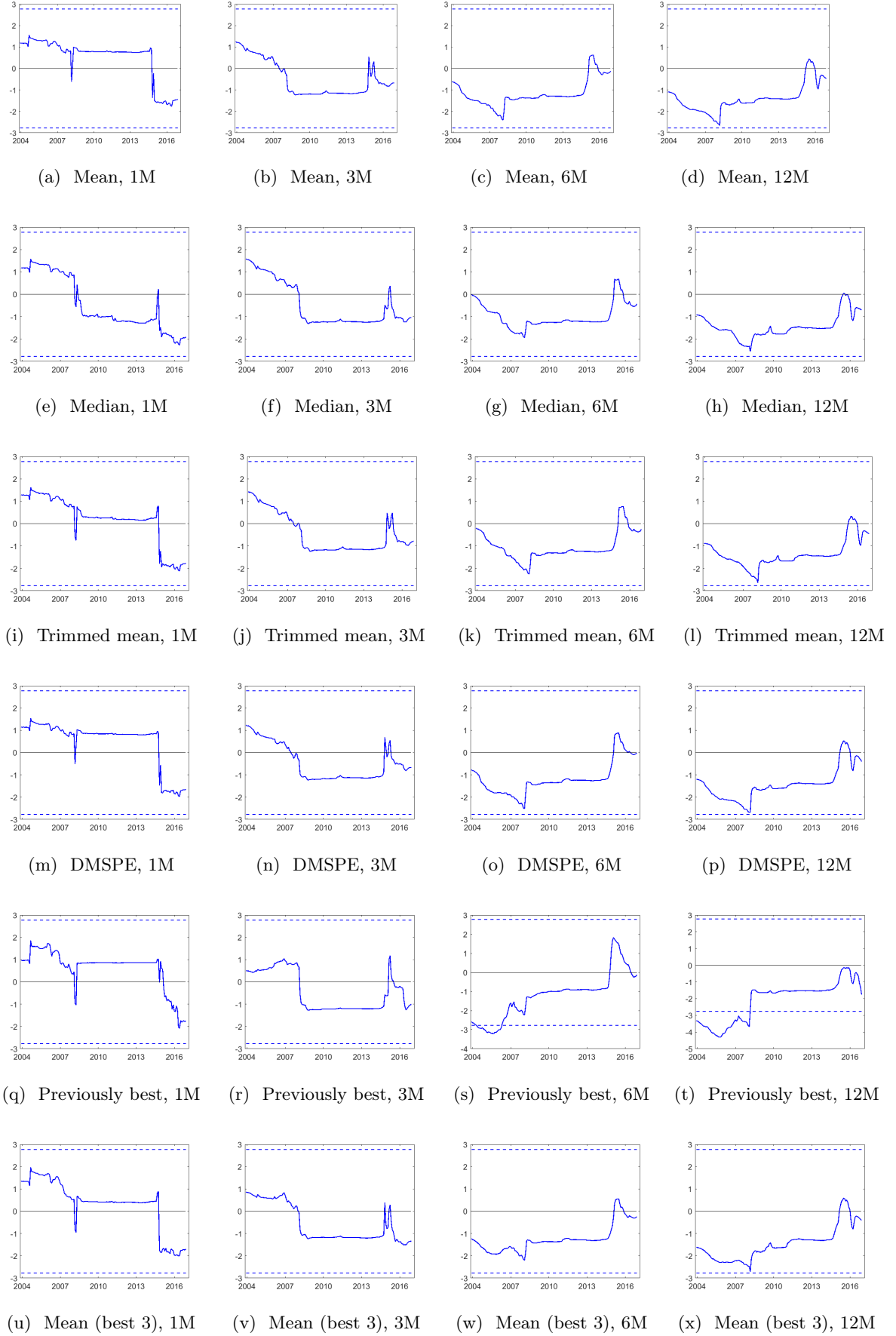


Figure A.9: Fluctuation test applied to forecast combinations of the individual GARCH-MIDAS models. Squared errors. Dashed lines represent 10% confidence bands. Benchmark: GJR-GARCH(1,1) model. DMSPE with $\eta = 0.5$. Note that the year on the x-axis marks the end of the rolling window period, over which the test statistics is calculated. $l = 5$, $m = 78$

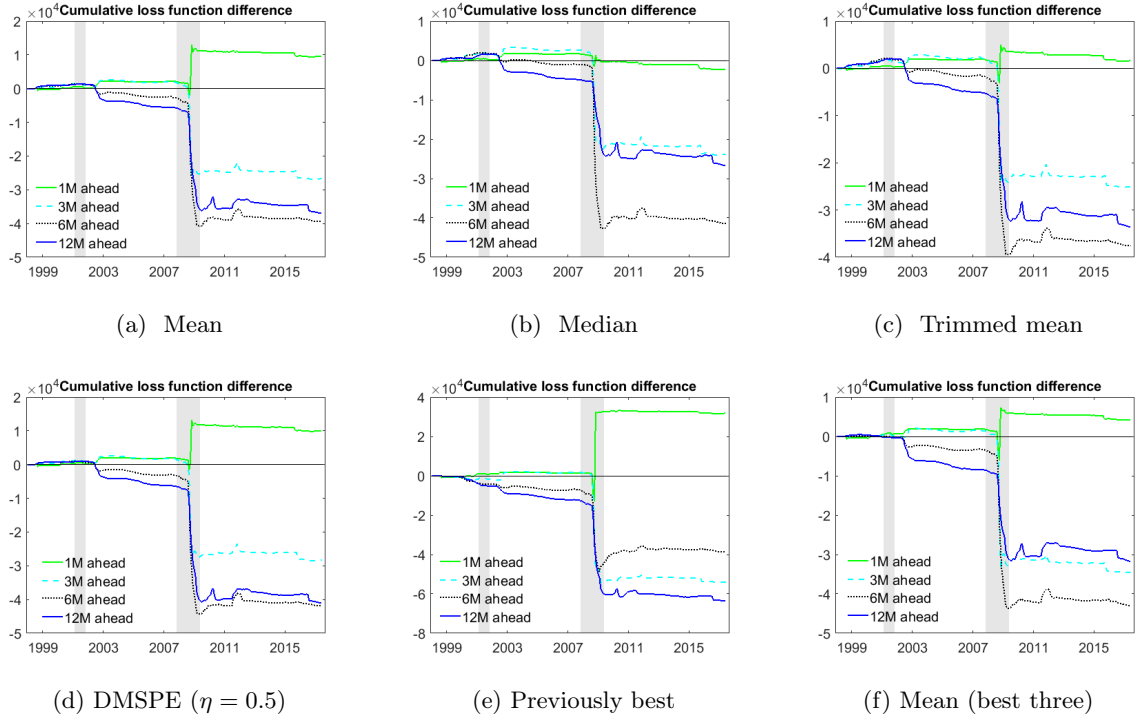


Figure A.10: Cumulative sum of loss function differences (squared errors) of forecast combinations of the individual GARCH-MIDAS models, compared to the GJR-GARCH(1,1) model ($(Loss_{combo} - Loss_{GARCH})^2$). An upward sloping segment thus indicates the GJR-GARCH model outperforms the GARCH-MIDAS model. Grey areas mark NBER dated US recessions.

Figure A.8 shows that the differences in forecasting performance were huge during the financial crisis, but qualitatively the results are similar to using absolute forecast errors. The main difference is that 12M ahead the improvements in GARCH-MIDAS forecasts seem to have been more modest than when using absolute errors.

Table A.5: Effect of business cycle (IP growth) on forecasting performance: MSFE ratios

	1 month ahead		3 month ahead		6 month ahead		12 month ahead	
	Pos. growth	Neg. growth	Positive	Negative	Positive	Negative	Positive	Negative
Buying Conditions index	1.00	0.97	1.01	0.89	1.03	0.90	1.07	0.96*
ISM New Orders index	1.00	0.99	1.03	0.99	1.05	0.99	0.99	0.99
Housing starts	1.01	0.96	1.03	0.95	1.05	0.91	1.01	0.88
ADS index	0.99	1.18	1.01	0.94	1.05	0.95	1.04	0.97
Term spread	1.03	1.30	1.01	1.00	1.00	0.93	0.96	0.90
Default spread	1.02	1.31	1.50	1.05	2.44	1.00	2.52	1.05
3M T-bill rate	1.04*	1.05	1.08**	1.03**	1.10**	1.03**	1.11**	1.01
Excess market return	0.98	1.51*	1.03	1.01	0.99	1.02	1.13	1.02
Realised volatility (RV)	1.02	1.33	1.56	1.22	2.85	1.18	3.69	1.11
First principal component	1.00	0.83	1.00	0.96	1.13	0.95	1.08	0.97
Second principal component	1.04	1.04	1.07	0.98	1.09	0.96	1.04	0.98
Third principal component	1.04	1.29	1.04	0.99	1.10	0.98	1.11	0.99

Benchmark: GJR-GARCH(1,1) mode. MSFE ratio: $\frac{MSFE_{GMX}}{MSFE_{GARCH}}$, where $MSFE_{GMX}$ stands for the mean squared forecast error from the GARCH-MIDAS model driven by some macroeconomic or financial data (X). A value below 1 means the GARCH-MIDAS model outperforms the GJR-GARCH(1,1) model. *, ** and *** indicate a rejection of the null hypothesis of equal (unconditional) predictive ability at the 10%, 5%, and 1% level, respectively, according to the Giacomini and White (2006) test. Low / high growth months are defined according to the sign of annualised industrial production growth data (manufacturing only, most recent release). $RV_t = \sum_{i=1}^{N_t} |r_{i,t}|$.

Table A.6: Effect of business cycle (NBER) on forecasting performance: MSFE ratios

	1 month ahead		3 month ahead		6 month ahead		12 month ahead	
	Expansion	Recession	Expansion	Recession	Expansion	Recession	Expansion	Recession
Buying Conditions index	1.00	0.96	1.01	0.88	1.00	0.89	1.02	0.96*
ISM New Orders index	1.00	0.99	1.03	0.99	1.03	0.99	0.97	1.00
Housing starts	1.00	0.96	1.03	0.94	1.03	0.89	0.99	0.87*
ADS index	0.98	1.21	1.01	0.94	1.02	0.94*	1.00	0.98
Term spread	1.05	1.33	1.02	1.00	0.97	0.93	0.91*	0.91*
Default spread	1.01	1.36	1.36	1.04	1.97	0.98	2.13	0.99
3M T-bill rate	1.03	1.05	1.08**	1.02**	1.09**	1.02**	1.08*	1.01***
Excess market return	1.01	1.59	1.01	1.01	0.98	1.03	1.08	1.03*
Realised volatility (RV)	1.02	1.39	1.37	1.25	2.19	1.19	3.19	0.97
First principal component	0.99	0.81	1.00	0.95	1.09	0.94	1.04	0.97
Second principal component	1.03	1.04	1.06	0.97	1.06	0.95	1.01	0.98
Third principal component	1.04	1.33	1.04	0.99	1.03	0.98	1.06	0.99*

Benchmark: GJR-GARCH(1,1) mode. MSFE ratio: $\frac{MSFE_{GMX}}{MSFE_{GARCH}}$, where $MSFE_{GMX}$ stands for the mean squared forecast error from the GARCH-MIDAS model driven by some macroeconomic or financial data (X). A value below 1 means the GARCH-MIDAS model outperforms the GJR-GARCH(1,1) model, and vice versa. *, ** and *** indicate a rejection of the null hypothesis of equal (unconditional) predictive ability at the 10%, 5%, and 1% level, respectively, according to the Giacomini and White (2006) test. Recession months are defined according to the NBER Business Cycle Dating Committee. $RV_t = \sum_{i=1}^{N_t} |r_{i,t}|$.

Table A.7: Effect of volatility environment on forecasting performance: MSFE ratios

	1 month ahead		3 month ahead		6 month ahead		12 month ahead	
	Low vola	High vola	Low vola	High vola	Low vola	High vola	Low vola	High vola
Buying Conditions index	0.98	0.98	0.72	0.91	0.71**	0.92	0.73	0.98
ISM New Orders index	0.92**	1.00	0.70	1.01	0.72**	1.00	0.64**	1.00
Housing starts	1.01	0.98	0.75	0.96	0.86	0.93	0.74	0.90
ADS index	0.98	1.13	0.83	0.96	0.87	0.97	0.83	0.98
Term spread	1.00	1.23	0.72	1.01	0.58***	0.94	0.37***	0.92
Default spread	0.98	1.23	0.90	1.12	1.03	1.22	3.66	1.21
3M T-bill rate	0.95**	1.05*	0.71*	1.04***	0.72***	1.04***	0.61***	1.04***
Excess market return	0.90*	1.37*	0.86	1.02	1.26*	1.01	1.56	1.03
Realised volatility (RV)	0.78**	1.26	0.72	1.29	0.97	1.44	3.80	1.44
First principal component	0.92*	0.88	0.81	0.97	0.93	0.97	1.00	0.98
Second principal component	0.91*	1.04	0.59**	1.00	0.51***	0.99	0.39***	1.00
Third principal component	0.97	1.22	0.69*	1.01	0.68**	1.00	0.53***	1.01

Benchmark: GJR-GARCH(1,1) model. MSFE ratio: $\frac{MSFE_{GMX}}{MSFE_{GARCH}}$, where $MSFE_{GMX}$ stands for the mean squared forecast error from the GARCH-MIDAS model driven by some macroeconomic or financial data (X). A value below 1 means the GARCH-MIDAS model outperforms the GJR-GARCH(1,1) model. *, ** and *** indicate a rejection of the null hypothesis of equal (unconditional) predictive ability at the 10%, 5%, and 1% level, respectively, according to the Giacomini and White (2006) test. High / low volatility months are defined according to the VIX index, where the median over the sample period is the cut off point. $RV_t = \sum_{i=1}^{N_t} |r_{i,t}|$.

Table A.8: Effect of financial market stress on forecasting performance: MSFE ratios

	1 month ahead		3 month ahead		6 month ahead		12 month ahead	
	Low stress	High stress	Low stress	High stress	Low stress	High stress	Low stress	High stress
Buying Conditions index	0.97	0.98	0.98	0.90	1.02	0.90	1.14	0.96*
ISM New Orders index	0.97	1.00	0.97	1.00	1.01	1.00	0.95	1.00
Housing starts	0.98	0.97	0.97	0.96	1.00	0.92	0.93	0.90
ADS index	0.96	1.16	0.96	0.96	1.01	0.96	0.90	0.98
Term spread	1.01	1.27	1.02	1.00	1.02	0.93	0.94	0.91
Default spread	0.97	1.29	0.98	1.14	1.83	1.14	3.40	1.01
3M T-bill rate	1.00	1.06*	1.00	1.04***	1.04	1.04***	1.04	1.03***
Excess market return	0.99	1.45*	0.99	1.02	0.95	1.03	1.15	1.02
Realised volatility (RV)	0.97	1.32	1.07	1.31	1.95	1.37	6.17	0.95**
First principal component	0.98	0.86	0.89*	0.97	1.01	0.97	1.08	0.97
Second principal component	1.00	1.05	1.05	0.99	1.08	0.97	0.98	0.99
Third principal component	1.03	1.26	1.03	0.99	1.05	0.99	1.13	0.99

Benchmark: GJR-GARCH(1,1) model. MSFE ratio: $\frac{MSFE_{GMX}}{MSFE_{GARCH}}$, where $MSFE_{GMX}$ stands for the mean squared forecast error from the GARCH-MIDAS model driven by some macroeconomic or financial data (X). A value below 1 means the GARCH-MIDAS model outperforms the GJR-GARCH(1,1) model, and vice versa. *, ** and *** indicate a rejection of the null hypothesis of equal (unconditional) predictive ability at the 10%, 5%, and 1% level, respectively, according to the Giacomini and White (2006) test. High / low financial stress months are defined according to the St. Louis Fed Financial Stress Index: 115 high stress months and 143 low stress months. $RV_t = \sum_{i=1}^{N_t} |r_{i,t}|$.

A.6 Robustness check: Estimation and weighting scheme

In this appendix I discuss the robustness of the results to (i) the weighting scheme, i.e. fixed weights instead of weights re-estimated each period, and (ii) the estimation scheme, i.e. expanding window instead of rolling window. Thus I have, first of all, estimated the models over the full sample, saved the weights of the weighting schemes, and then re-estimated the models using a rolling window with the weights fixed at the full-sample weights. The other parameters of the GARCH-MIDAS model are re-estimated each period. Secondly, I have estimated each GARCH-MIDAS model using an expanding window, i.e., adding one month to the estimation in each period. The differences in the forecasts produced, the in-sample fit (in terms of the variance ratio) and the parameter estimates are discussed below.

Table A.9: Full sample MAFE ratios: fixed vs. re-estimated weights

	1M ahead	2M ahead	3M ahead	6M ahead	9M ahead	12M ahead
Buying Conditions index	0.99	0.99	0.99	1.00	0.99	1.00
ISM New Orders index	1.00	1.00	0.99	1.01	1.01	1.01
Housing starts	1.01	1.01	0.99	1.00	1.03	1.03
ADS index	0.97	1.02	1.02	1.04	1.03	1.02
Term spread	1.00	1.00	0.99	1.00	1.01	1.01
Default spread	1.00	1.00	0.99	1.00	1.01	1.00
3M T-bill rate	1.00	0.99	0.99	0.99	0.99	0.99
Excess market return	0.93	0.96	0.98	0.96	0.94	0.92
Realised volatility (RV)	1.00	1.00	1.01	1.01	1.01	1.00
First principal component	1.00	1.01	1.01	1.01	1.01	1.01
Second principal component	0.99	1.00	1.00	1.00	1.00	1.00
Third principal component	0.95	0.98	0.98	0.98	0.97	0.95

MAFE ratio: $\frac{MAFE_{GMXfix}}{MAFE_{GMX}}$, where $MAFE_{GMXfix}$ ($MAFE_{GMX}$) stands for the mean absolute forecast error from the GARCH-MIDAS model driven by some macroeconomic or financial data (X) estimated using fixed (re-estimated) weights. A value below 1 means the fixed weights forecast outperforms the forecast using weights re-estimated in each period. $RV_t = \sum_{i=1}^{N_t} |r_{i,t}|$.

Starting with (i), over the full sample the choice of fixed or rolling weights has little effect on the out-of-sample forecasts (Table A.9), with the exception of the excess market return and PC3, for which fixed weights produce a more accurate forecast. Figure A.13 looks at the cumulative loss function differences vis-à-vis the GJR-GARCH(1,1) model. Mostly the weighting scheme does not matter much for the relative performance of the models over time. However, large difference occur for the GARCH-MIDAS model driven by the ADS index, housing starts, excess market returns and PC3. For the ADS index and housing starts fixing the weight(s) lead to a clearly different performance during/after the financial crisis: the 12M (and also 3M for the ADS index) ahead forecasts are worse when fixing the weights, while for the ADS index the 1M ahead

Table A.10: Full sample MSFE ratios: fixed vs. re-estimated weights

	1M ahead	2M ahead	3M ahead	6M ahead	9M ahead	12M ahead
Buying Conditions index	0.99	1.01	1.02	1.01	1.00	1.00
ISM New Orders index	0.99	0.99	0.99	0.97	1.00	1.00
Housing starts	1.01	1.01	1.00	1.02	1.03	1.05
ADS index	0.80	1.03	1.03	1.03	1.02	1.02
Term spread	0.99	1.00	1.00	1.00	1.00	1.00
Default spread	1.05	1.00	0.98	1.00	1.01	1.01
3M T-bill rate	1.00	1.00	1.00	1.00	1.00	1.00
Excess market return	0.81	0.96	0.98	0.98	0.95	0.96
Realised volatility (RV)	1.11	1.08	1.08	1.11	1.11	1.12
First principal component	1.01	1.01	1.01	1.01	1.01	1.01
Second principal component	0.99	1.00	1.00	1.01	1.00	0.99
Third principal component	0.83	0.98	1.01	1.00	0.98	0.98

MSFE ratio: $\frac{MSFE_{GMXfix}}{MSFE_{GMX}}$, where $MSFE_{GMXfix}$ ($MSFE_{GMX}$) stands for the mean squared forecast error from the GARCH-MIDAS model driven by some macroeconomic or financial data (X) estimated using fixed (re-estimated) weights. A value below 1 means the fixed weights forecast outperforms the forecast using weights re-estimated in each period. $RV_t = \sum_{i=1}^{N_t} |r_{i,t}|$.

forecast is better. For excess returns and PC3, in particular 12 months ahead, the forecasts using fixed weights perform clearly better than the forecasts from the models where the weights are re-estimated each period.

Mostly the differences in in-sample fit (variance ratios) are relatively small (Figure A.14), with the exception of the GARCH-MIDAS model driven by the term spread towards the end of the period and for the model driven by excess market returns. In both cases the model with weight parameters re-estimated each period produces a better fit.

Figure A.15 shows the estimates for θ for fixed and re-estimated weights. There are mostly relatively small differences especially in those periods when the weighting schemes differ from each other the most. The main exception is again excess market returns, for which θ has the opposite sign when weights are fixed, compared to other estimation schemes. However, as we can see from Figure A.16 the negative θ estimate from the fixed weights model is only borderline statistically significant. The change in sign of the second and third PC still holds, confirming it is not a consequence of imprecisely estimated weights but rather the changing composition of the PC.

Moving on to (ii), the expanding window estimation scheme leads to lower forecast errors in most cases, which is especially pronounced for the MAFE ratios. Note that the statistical tests by Giacomini and White (2006) and Giacomini and Rossi (2010) are not valid for the expanding window. Using a rolling window estimation scheme does, however, seem to underestimate the

Table A.11: Full sample MAFE ratios: expanding vs. rolling window estimation scheme

	1M ahead	2M ahead	3M ahead	6M ahead	9M ahead	12M ahead
GJR-GARCH(1,1)	0.98	0.98	0.97	0.95	0.97	0.97
Buying Conditions index	0.98	0.99	0.99	0.98	0.97	0.96
ISM New Orders index	0.98	0.98	0.96	0.96	0.98	0.97
Housing starts	0.99	0.98	0.95	0.96	0.98	0.98
ADS index	0.95	0.99	0.97	0.98	0.98	0.98
Term spread	0.97	0.97	0.97	1.00	1.01	1.01
Default spread	0.90	0.88	0.85	0.82	0.83	0.83
3M T-bill rate	0.97	0.96	0.95	0.95	0.96	0.96
Excess market return	0.96	0.97	0.96	0.94	0.95	0.94
Realised volatility (RV)	0.96	0.95	0.95	0.96	0.96	0.96
First principal component	0.99	0.99	0.97	0.96	0.95	0.96
Second principal component	0.97	0.99	0.98	0.99	1.00	1.00
Third principal component	0.94	0.97	0.96	0.96	0.96	0.95

MAFE ratio: $\frac{MAFE_{GMXexp}}{MAFE_{GMX}}$, where $MAFE_{GMXexp}$ ($MAFE_{GMX}$) stands for the mean absolute forecast error from the GARCH-MIDAS model driven by some macroeconomic or financial data (X) estimated using expanding (rolling) estimation scheme. A value below 1 means the expanding window forecast outperforms the rolling window forecast. $RV_t = \sum_{i=1}^{N_t} |r_{i,t}|$.

Table A.12: Full sample MSFE ratios: expanding vs. rolling window estimation scheme

	1M ahead	2M ahead	3M ahead	6M ahead	9M ahead	12M ahead
GJR-GARCH(1,1)	1.03	0.99	0.98	0.99	1.00	1.00
Buying Conditions index	1.09	1.07	1.08	1.06	1.03	1.01
ISM New Orders index	1.04	0.99	0.99	1.00	1.00	1.00
Housing starts	1.09	1.03	1.02	1.04	1.05	1.04
ADS index	0.87	1.06	1.03	1.02	1.02	1.01
Term spread	0.98	0.99	1.02	1.04	1.04	1.04
Default spread	0.87	0.90	0.88	0.83	0.82	0.81
3M T-bill rate	1.02	0.99	0.98	0.99	1.00	1.00
Excess market return	0.90	0.96	0.97	0.98	0.98	0.97
Realised volatility (RV)	0.98	0.95	0.96	0.96	0.96	0.96
First principal component	1.12	1.03	1.01	1.01	1.00	1.00
Second principal component	1.05	1.03	1.02	1.04	1.03	1.02
Third principal component	0.89	0.99	1.02	1.01	1.00	1.00

MSFE ratio: $\frac{MSFE_{GMXexp}}{MSFE_{GMX}}$, where $MSFE_{GMXexp}$ ($MSFE_{GMX}$) stands for the mean squared forecast error from the GARCH-MIDAS model driven by some macroeconomic or financial data (X) estimated using expanding (rolling) estimation scheme. A value below 1 means the expanding window forecast outperforms the rolling window forecast. $RV_t = \sum_{i=1}^{N_t} |r_{i,t}|$.

benefits of using macroeconomic and financial data to forecast stock market volatility, compared to an expanding window estimation scheme. As expected, the expanding window leads to more stable parameter estimates which are closer to the full-sample estimates for all models (Figures A.15 and A.17), and as seen from Tables ?? and ?? this also has a favourable impact on many forecasts. Figure A.16 indicates that the estimate of θ also tend to be more strongly

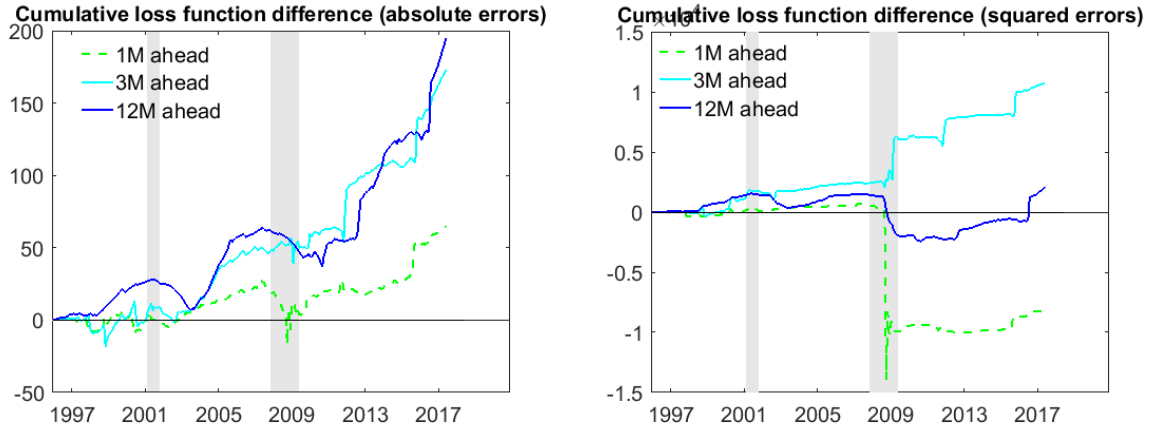


Figure A.11: Cumulative loss function difference between the GJR-GARCH model, estimated using either a rolling window or an expanding window. When the line is upward sloping the model estimated using the expanding window outperforms the model estimated using a rolling window.

statistically significant for the expanding window estimation scheme. From Figure A.14 we can see that also the variance ratios fluctuate less when an expanding scheme is used, but they tend to be lower, implying a worse in-sample fit.

Regarding the cumulative sum of loss function differences, there is a significant difference already for the benchmark GJR-GARCH model (Figure A.11), with the expanding window scheme performing better. In particular, the GJR-GARCH model estimated using the expanding window performs better on all horizons and in most time periods. Secondly, when comparing the GARCH-MIDAS models to the GJR-GARCH model we see that the largest differences, in favour of the expanding window scheme, occur for the ADS index (1M horizon), the default spread (all horizons) and the 3M T-bill rate (3M horizon). On the 12 month horizon the expanding window estimation scheme leads to less accurate forecasts (relative to the benchmark) for example when the GARCH-MIDAS model is driven by the term spread, housing starts or the second principal component.

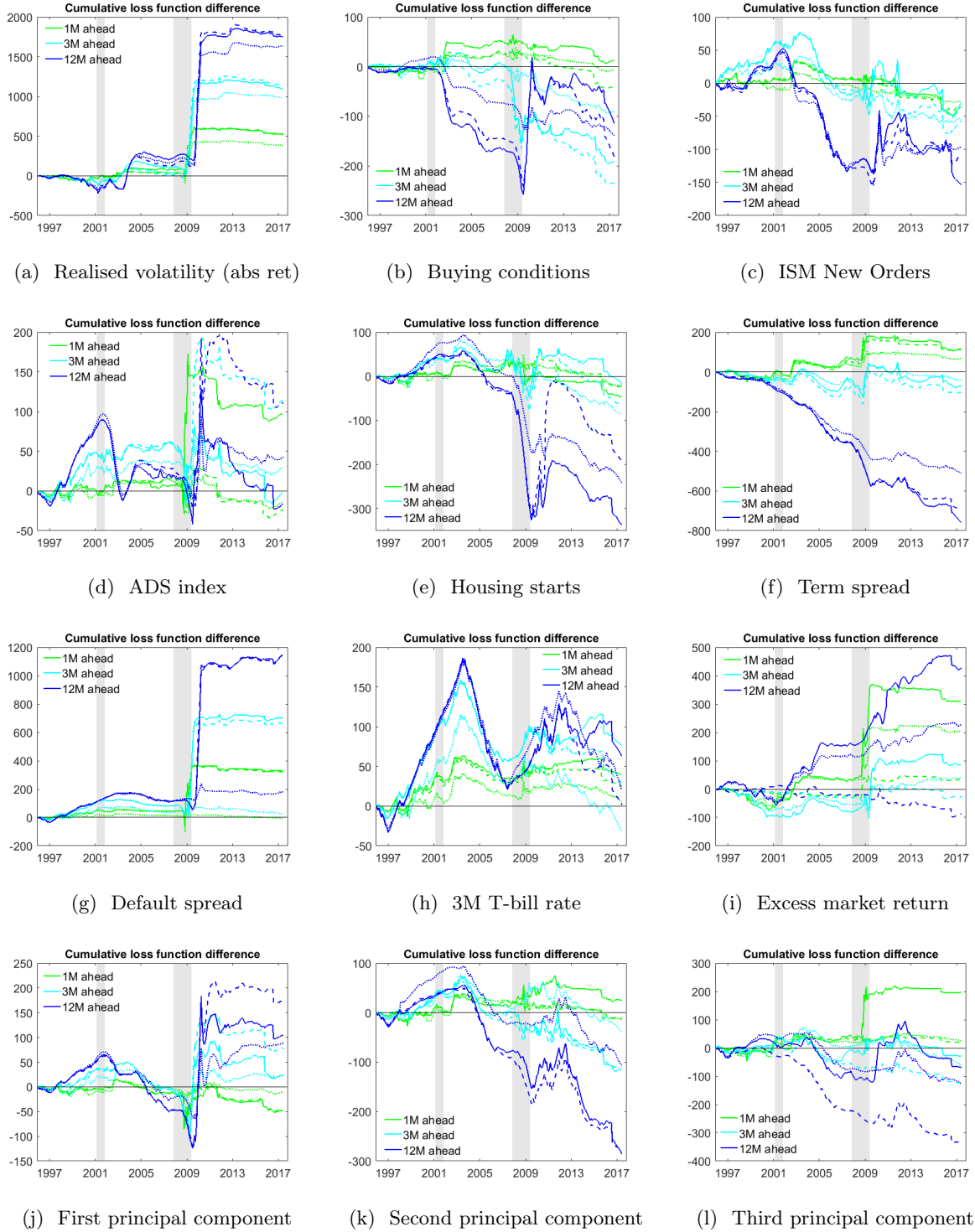
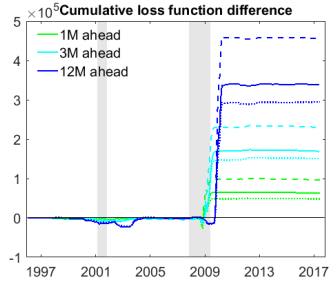
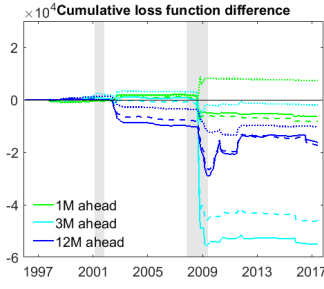


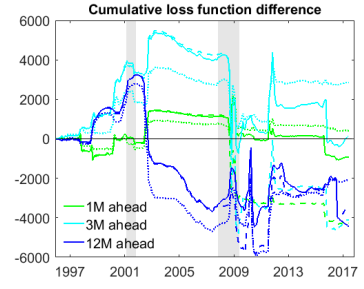
Figure A.12: Cumulative sum of loss function differences (absolute errors) of rolling window GARCH-MIDAS models with fixed weights (dashed line), GARCH-MIDAS models estimated using an expanding window (dotted line), and rolling window GARCH-MIDAS models with weights re-estimated each period (solid line). Baseline model: the GJR-GARCH(1,1) model, estimated using either a rolling window or an expanding window. When the line is upward sloping the GJR-GARCH model outperforms the GARCH-MIDAS model.



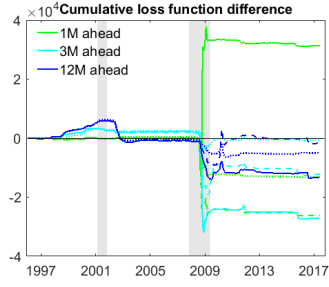
(a) Realised volatility (abs ret)



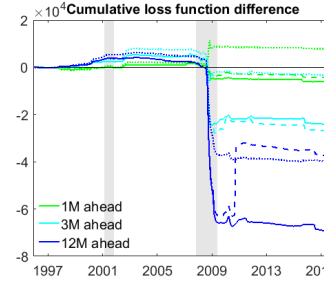
(b) Buying conditions



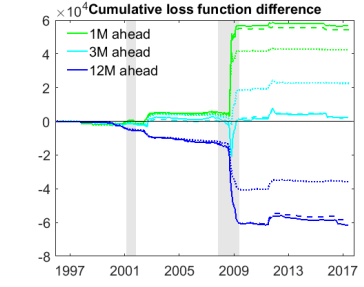
(c) ISM New Orders



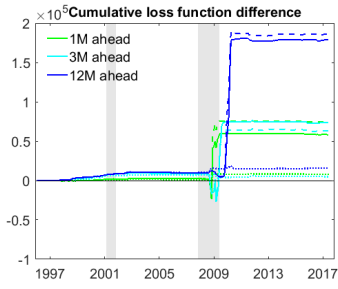
(d) ADS index



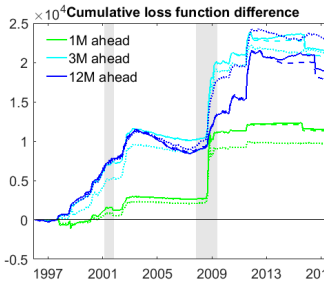
(e) Housing starts



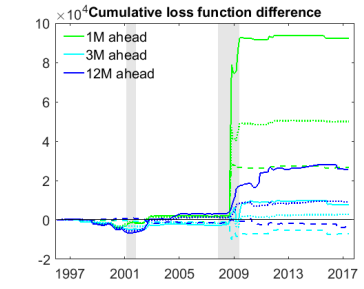
(f) Term spread



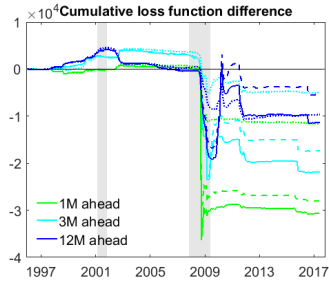
(g) Default spread



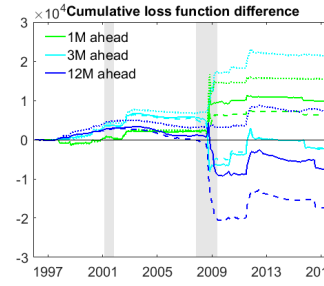
(h) 3M T-bill rate



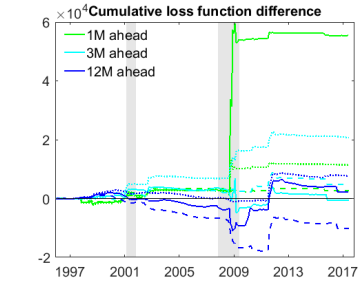
(i) Excess market return



(j) First principal component



(k) Second principal component



(l) Third principal component

Figure A.13: Cumulative sum of loss function differences (squared errors) of rolling window GARCH-MIDAS models with fixed weights (dashed line), GARCH-MIDAS models estimated using an expanding window (dotted line), and rolling window GARCH-MIDAS models with weights re-estimated each period (solid line). Baseline model: the GJR-GARCH(1,1) model, estimated using either a rolling window or an expanding window. When the line is upward sloping the GJR-GARCH model outperforms the GARCH-MIDAS model.

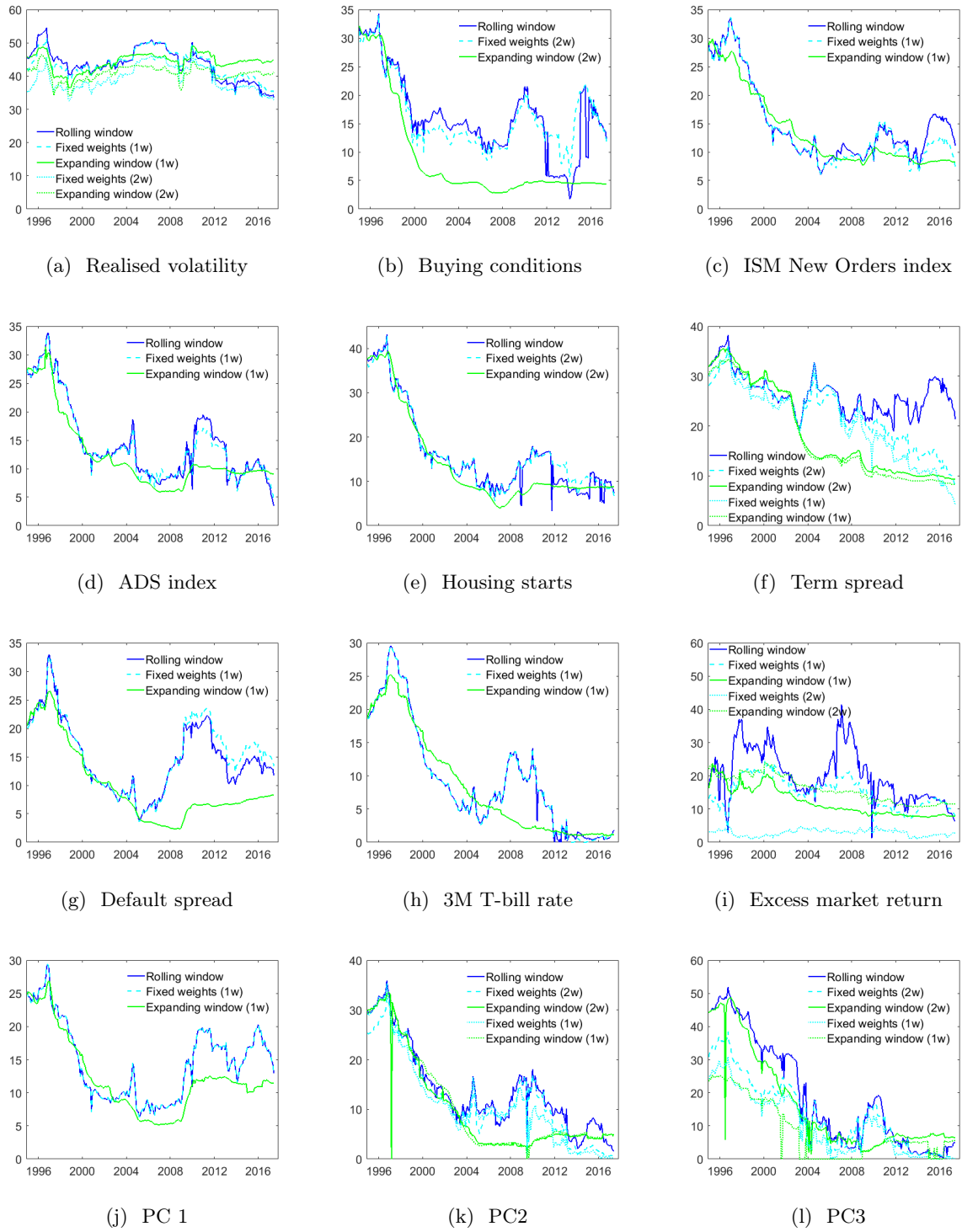


Figure A.14: Variance ratios of the rolling window GARCH-MIDAS models with fixed weights and the rolling and expanding window GARCH-MIDAS models with the weights re-estimated each period.

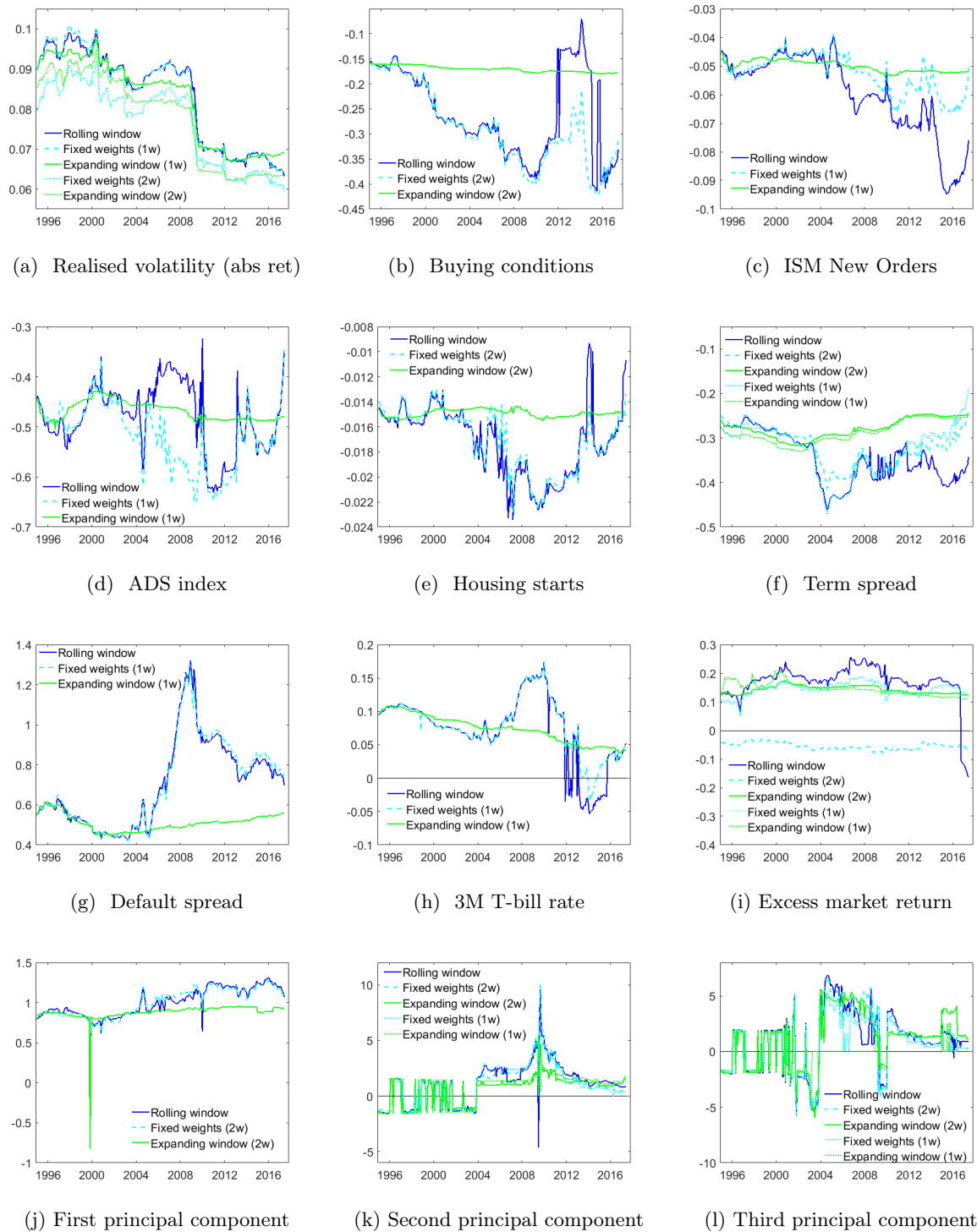


Figure A.15: Estimates for θ of the rolling window GARCH-MIDAS models with fixed weights and the rolling and expanding window GARCH-MIDAS models with the weights re-estimated each period.

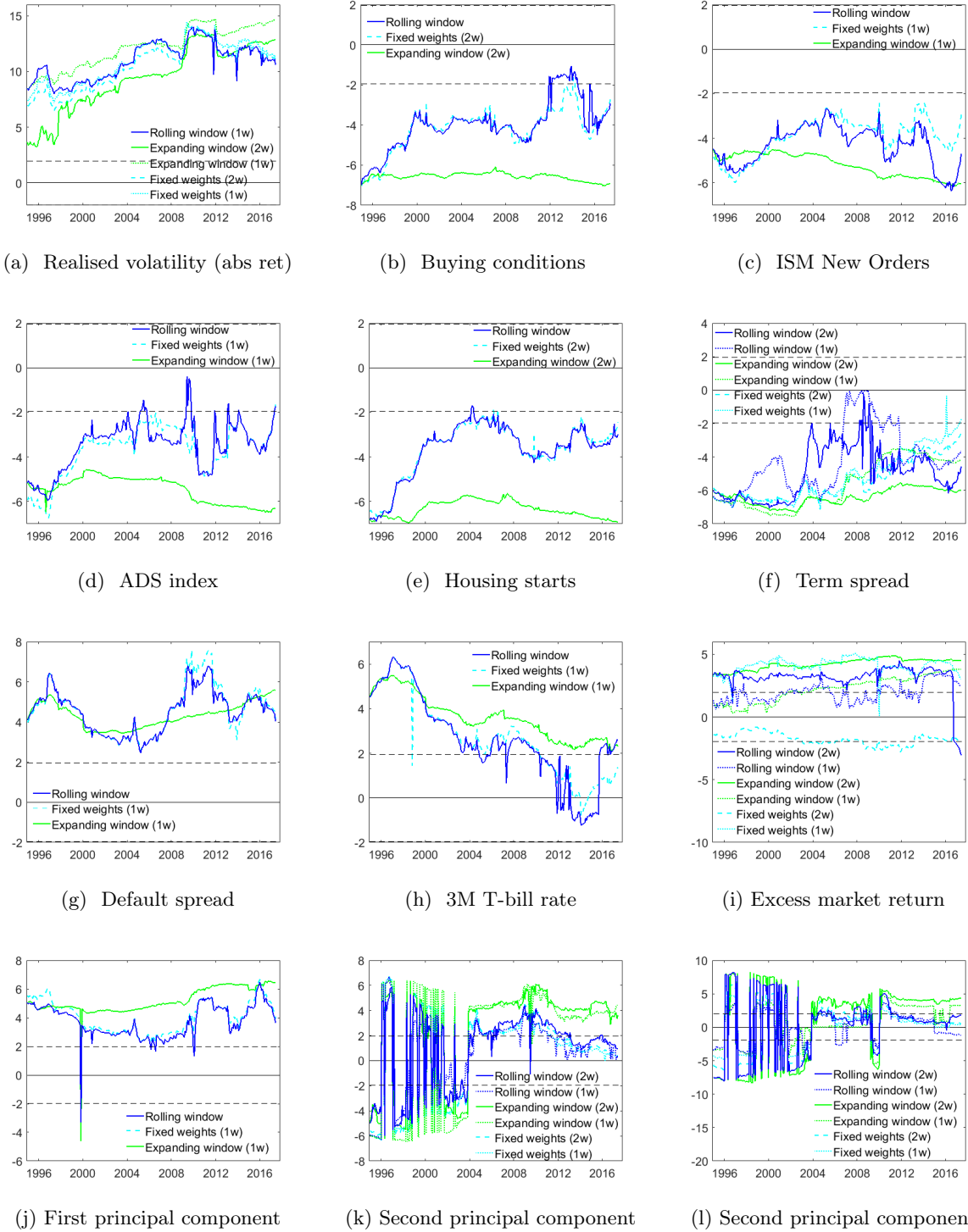


Figure A.16: t-statistics for the estimated θ parameters of the rolling window GARCH-MIDAS models with fixed weights and the rolling and expanding window GARCH-MIDAS models with the weights re-estimated each period.

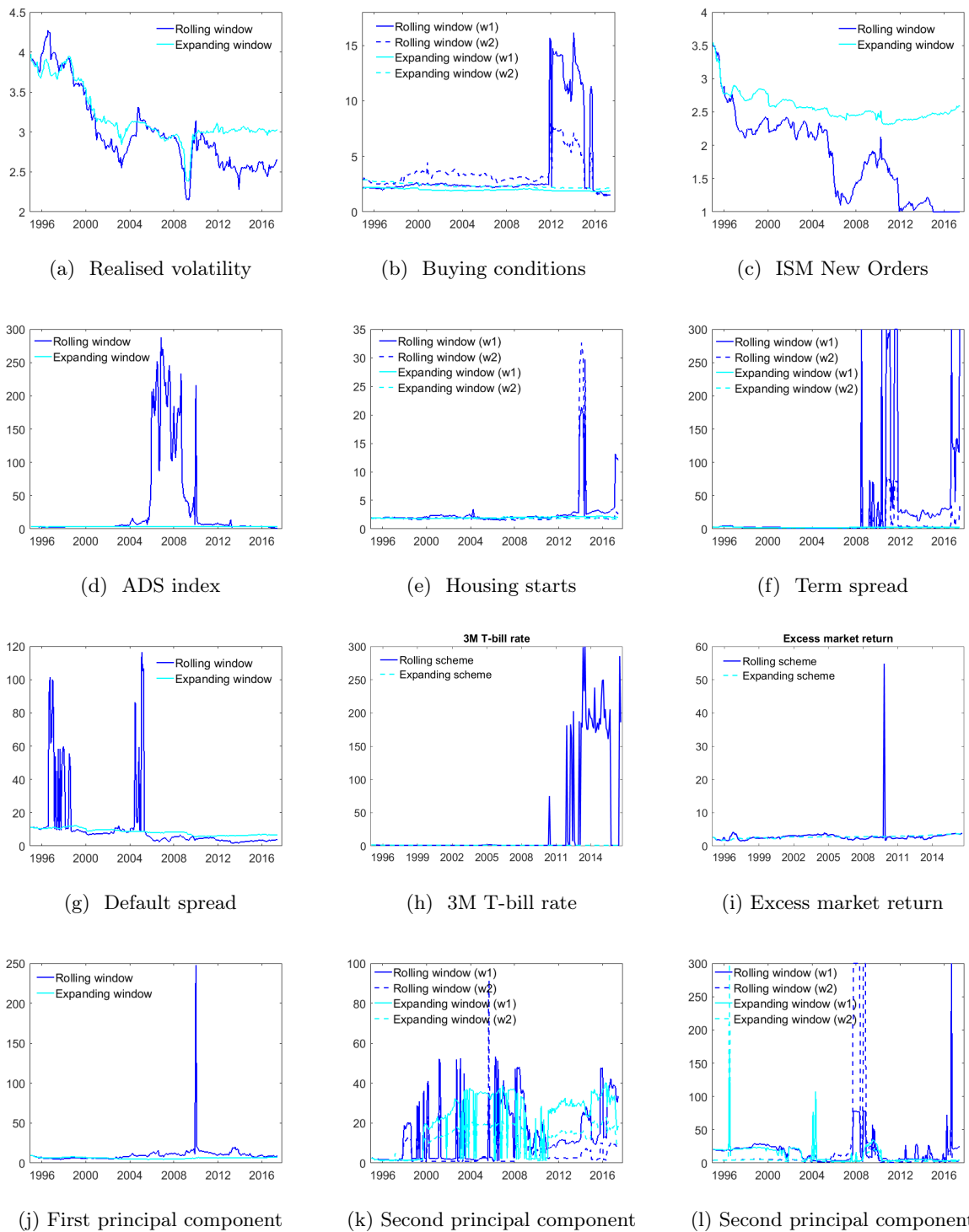


Figure A.17: Estimates for w of the GARCH-MIDAS models estimated using a rolling window (darker line) and GARCH-MIDAS models estimated using an expanding window (lighter line).

A.7 Additional time-varying forecasting results

This appendix presents the cumulative sums of the loss function differences and the graphical results of the Fluctuation test, complementing those presented in Section 6.2. The decision to exclude these figures from the main text relies on 1) the 3M T-bill rate leads to a mainly similar Fluctuation test result as the default spread, 2) the excess market return and realised volatility lead to a generally weak performance throughout the sample period, as was clear from the full-sample results, making the time-varying results less interesting, and 3) the principal components driven models lead to largely similar, and at least no better, forecast accuracy as the series they are based on.

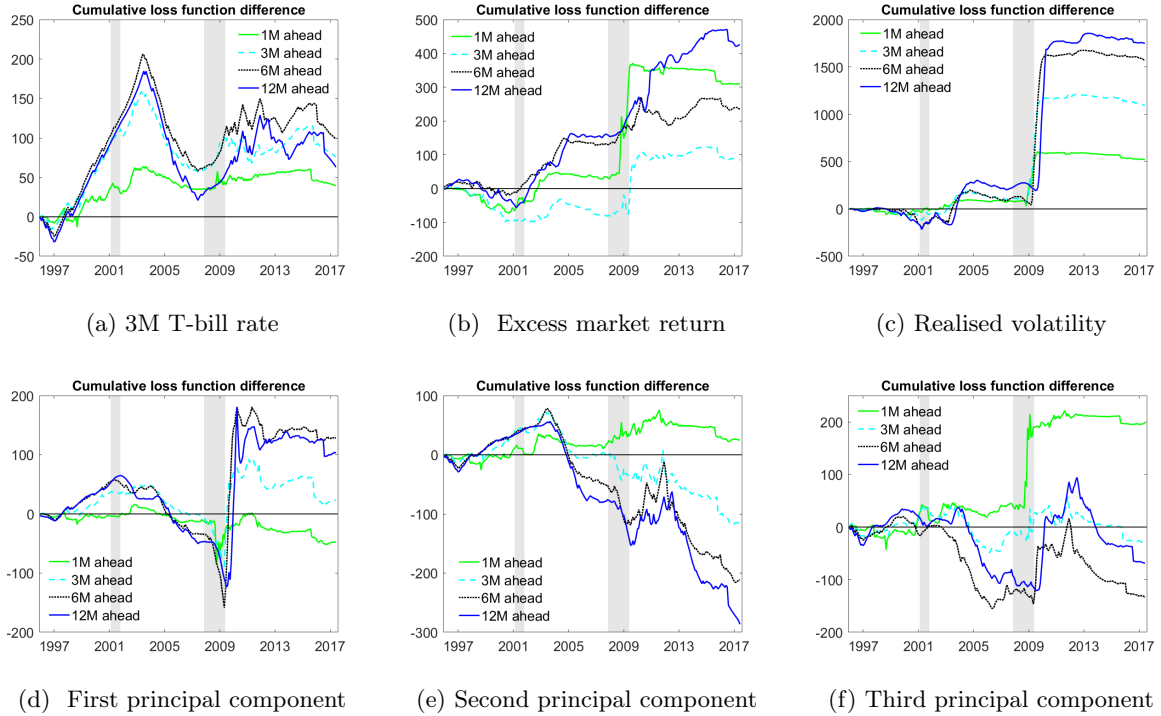


Figure A.18: Cumulative sum of loss function differences (absolute errors) ($|Loss_{GMX}| - |Loss_{GARCH}|$). An upward sloping segment thus indicates the GJR-GARCH model outperforms the GARCH-MIDAS model. Grey areas mark NBER dated US recessions.

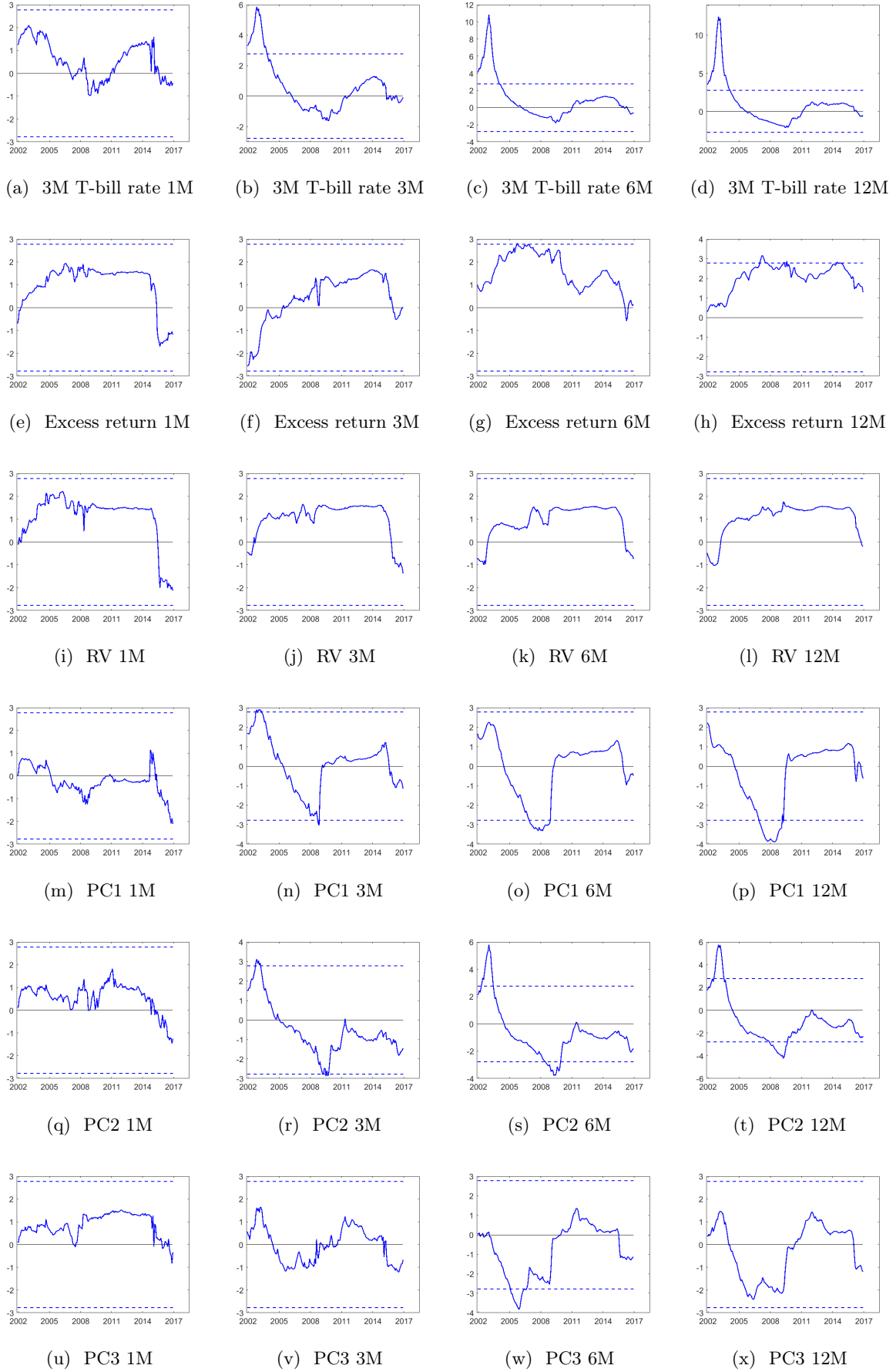
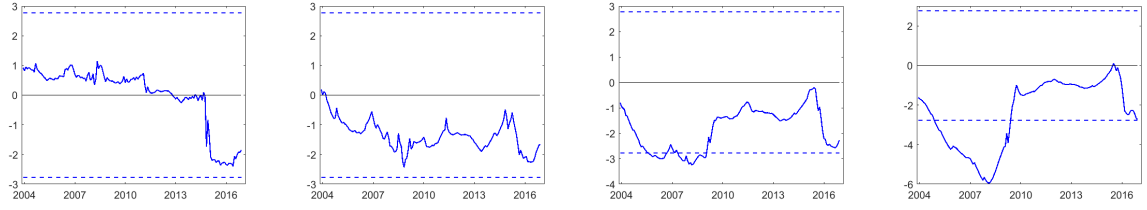
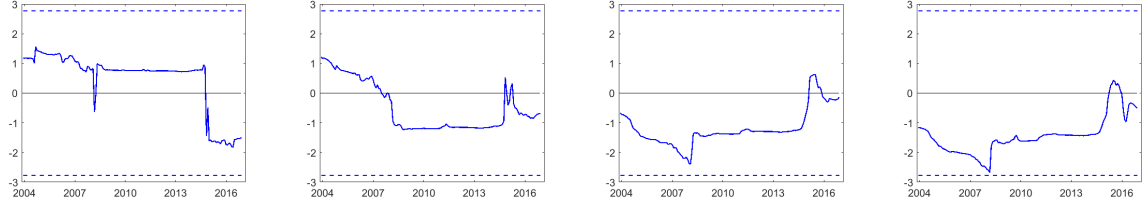


Figure A.19: Fluctuation test result for selected loss function differences. Dashed lines represent 10% confidence bands. Benchmark: GJR-GARCH(1,1) model. Note that the year on the x-axis marks the end of the rolling window period, over which the test statistics is calculated. $l = 5$, $m = 78$

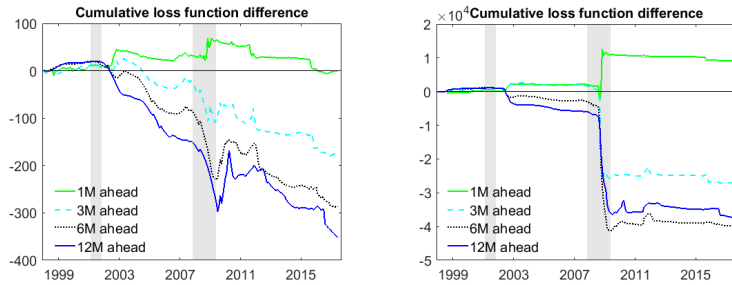


(a) DMAPE ($\eta = 1$), 1M (b) DMAPE ($\eta = 1$), 3M (c) DMAPE ($\eta = 1$), 6M (d) DMAPE ($\eta = 1$), 12M



(e) DMSPE ($\eta = 1$), 1M (f) DMSPE ($\eta = 1$), 3M (g) DMSPE ($\eta = 1$), 6M (h) DMSPE ($\eta = 1$), 12M

Figure A.20: Fluctuation test applied to forecast combinations of the individual GARCH-MIDAS models. First row is for absolute errors and second row for squared errors. Dashed lines represent 10% confidence bands. Benchmark: GJR-GARCH(1,1) model. Note that the year on the x-axis marks the end of the rolling window period, over which the test statistics is calculated. $l = 5$, $m = 78$



(a) DMAPE ($\eta = 1$)

(b) DMSPE ($\eta = 1$)

Figure A.21: Cumulative sum of loss function differences of forecast combinations of the individual GARCH-MIDAS models, compared to the GJR-GARCH(1,1) model ($|Loss_{combo} - Loss_{GARCH}|$). Panel (a) presents results for absolute errors and Panel (b) for squared errors. An upward sloping segment thus indicates the GJR-GARCH model outperforms the GARCH-MIDAS model. Grey areas mark NBER dated US recessions.

A.8 Robustness check: Effect of economic environment

I begin by plotting the NBER recession dates, industrial production growth, VIX index and the St. Louis Fed Financial Stress Index in Figures A.22 and A.23, to illustrate how the data is divided.

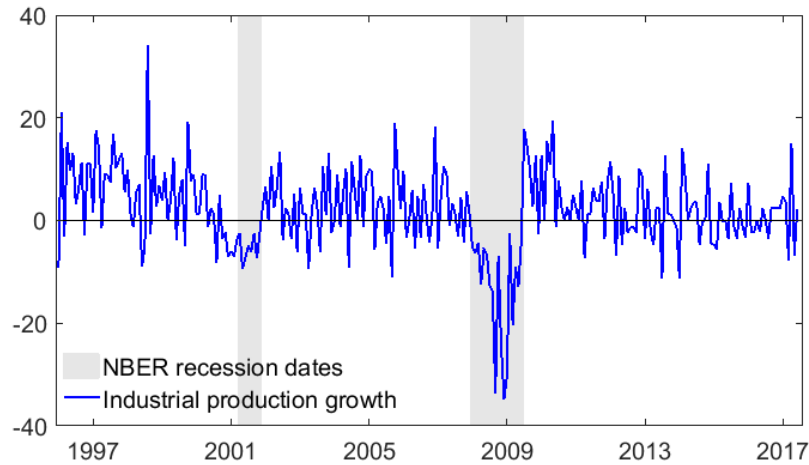


Figure A.22: NBER recession dates and industrial production growth. Zero is the cut-off point for industrial production growth.

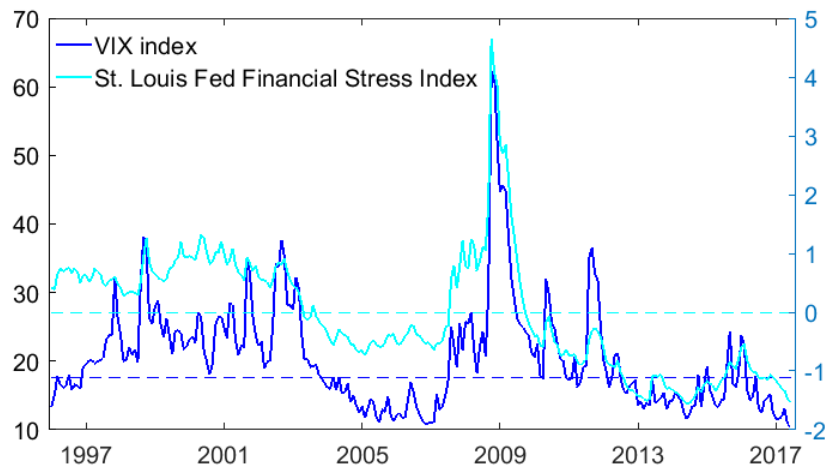


Figure A.23: VIX index and St. Louis Fed Financial Stress Index. Dashed lines denote the cut-off point of high versus low volatility (or financial stress) periods for each series.

There has only been two recessions during the sample period (from March 2001 to November 2001 (8 months) and from December 2007 to June 2009 (18 months)) and two longer episodes of negative industrial production growth, but several shorter spells of negative growth. The VIX index divides the out-of-sample period into roughly four episodes when using the median as the cut-off: high volatility from 1996 to 2003, low volatility from 2003 until the beginning of the

financial crisis in 2007, the financial crisis and its aftermath, and the largely low volatility since then. The St. Louis Fed Financial Stress Index, which is defined so that zero is the cut-off point between high and low stress regimes, divides the data roughly similarly.

Table A.13: Effect of business cycle (NBER) on forecasting performance: MAFE ratios

	1 month ahead		3 month ahead		6 month ahead		12 month ahead	
	Expansion	Recession	Expansion	Recession	Expansion	Recession	Expansion	Recession
Buying Conditions index	1.00	1.01	0.99	0.91**	0.99	0.90*	0.99	0.96
ISM New Orders index	0.99	1.00	0.99	0.99	0.99	0.99*	0.95	1.01
Housing starts	1.00	0.96	1.02	0.95	1.01	0.91**	0.98	0.88**
ADS index	0.98	1.13	1.01	0.99	1.02	0.94**	1.01	0.98
Term spread	1.01	1.09	0.96	1.04	0.90***	0.93**	0.85***	0.90***
Default spread	1.02	1.26	1.13	1.16	1.28	1.05	1.32	0.99
3M T-bill rate	1.02	1.00	1.01	1.03***	1.01	1.03***	1.00	1.02***
Excess market return	0.99	1.32**	1.01	1.03	1.04	1.04***	1.09**	1.04***
Realised volatility (RV)	1.02	1.42*	1.13	1.40*	1.30	1.26	1.48	1.01
Principal component 1	0.99	0.99	1.02	0.98	1.06	0.95**	1.05	0.97
Principal component 2	1.01	1.00	0.98	0.98	0.95	0.97	0.94*	0.97
Principal component 3	1.02	1.12	0.98	1.02	0.96	1.00	0.99	0.98**

Benchmark: GJR-GARCH(1,1). MAFE ratio: $\frac{MAFE_{GMX}}{MAFE_{GARCH}}$, where $MAFE_{GMX}$ is the mean absolute forecast error of the GARCH-MIDAS model driven by some economic data (X). A value below 1 means the GARCH-MIDAS model outperforms the GJR-GARCH(1,1) model. *, ** and *** indicate a rejection of the null hypothesis of equal (unconditional) predictive ability at the 10%, 5%, and 1% level, respectively, according to the Giacomini and White (2006) test. Recession months are defined according to the NBER Business Cycle Dating Committee: 26 recession months and 232 expansion months. $RV_t = \sum_{i=1}^{N_t} |r_{i,t}|$.

Table A.14: Effect of financial market stress on forecasting performance (STLFISI): MAFE ratios

	1 month ahead		3 month ahead		6 month ahead		12 month ahead	
	Low stress	High stress	Low stress	High stress	Low stress	High stress	Low stress	High stress
Buying Conditions index	0.98	1.02	0.94**	0.98	0.95	0.96	1.02	0.96
ISM New Orders index	0.95**	1.01	0.92**	1.04**	0.92	1.02*	0.92	1.00
Housing starts	0.97	1.00	0.97	1.01	0.95	0.99	0.94	0.94
ADS index	0.96	1.06	0.96	1.02	0.98	1.00	0.99	1.00
Term spread	0.97	1.07	0.92	1.02	0.86**	0.94**	0.79***	0.90***
Default spread	0.96*	1.16	0.99	1.22	1.21	1.19	1.55	1.04*
3M T-bill rate	0.98	1.03*	0.94**	1.06***	0.93*	1.06***	0.92	1.05***
Excess market return	0.95	1.16**	1.00	1.02	1.06	1.03**	1.16**	1.03
Realised volatility (RV)	0.96	1.25**	1.04	1.31*	1.31	1.27	2.01	0.97
Principal component 1	0.96	1.00	0.95	1.03	1.02	1.03	1.07	0.99
Principal component 2	0.97	1.03	0.92*	1.01	0.88**	1.00	0.85**	1.00
Principal component 3	1.02	1.08	0.95	1.02	0.90**	1.01	0.97	1.00

Benchmark: GJR-GARCH(1,1). MAFE ratio: $\frac{MAFE_{GMX}}{MAFE_{GARCH}}$, where $MAFE_{GMX}$ is the mean absolute forecast error of the GARCH-MIDAS model driven by some economic data (X). A value below 1 means GARCH-MIDAS model outperforms the GJR-GARCH(1,1) model. *, ** and *** indicate rejection of the null hypothesis of equal (unconditional) predictive ability at the 10%, 5%, and 1% level, respectively, according to the Giacomini and White (2006) test. High / low financial stress months are defined according to the St. Louis Fed Financial Stress Index: 115 high and 143 low financial stress months. $RV_t = \sum_{i=1}^{N_t} |r_{i,t}|$

A.9 Additional results: Conditional predictive ability

This appendix presents additional results relating to Section 6.4, considering whether the forecast errors are predictable using annualised industrial production growth, NBER recessions dates, the financial market stress indicator (STLFSI) and past forecast loss differences.

Table A.15: Conditional test using NBER recession dates: MAFE ratios

	1M ahead	2M ahead	3M ahead	6M ahead	9M ahead	12M ahead
Buying Conditions	1.00	0.98	0.96	0.96*	0.97*	0.98*
ISM New Orders	0.99	1.00	0.99	0.99	0.98	0.97
Housing starts	0.99	0.99	1.00	0.97	0.95	0.94
ADS index	1.03	1.00	1.00	0.99	1.00	1.00
Term spread	1.03	1.03	0.99	0.91***	0.88***	0.87***
Default spread	1.09	1.13	1.14	1.20	1.22	1.20
3M T-bill rate	1.01	1.02	1.02	1.02	1.02	1.01
Excess market return	1.08	1.03	1.02	1.04*	1.06**	1.08**
Realised volatility	1.14	1.23*	1.22	1.29	1.32	1.31
First principal component	0.99	1.00	1.00	1.02	1.04	1.02
Second principal component	1.01	0.99	0.98	0.96	0.96	0.95*
Third principal component	1.05	1.00	0.99	0.98**	0.98*	0.99

Benchmark: $GJR-GARCH(1,1)$. MAFE ratio: $\frac{MAFE_{GMX}}{MAFE_{GARCH}}$, where $MAFE_{GMX}$ stands for the mean absolute forecast error from the GARCH-MIDAS model driven by some macroeconomic or financial data (X). A value below 1 means the GARCH-MIDAS model outperforms the $GJR-GARCH(1,1)$ model. *, ** and *** indicate a rejection of the null hypothesis of equal (conditional) predictive ability at the 10%, 5%, and 1% level, respectively, according to the Giacomini and White (2006) test. Test function: $h_t = [1 \ v_t]$, where v_t is the conditioning information. $RV_t = \sum_{i=1}^{N_t} |r_{i,t}|$.

Table A.16: Conditional test using industrial production data: MAFE ratios

	1M ahead	2M ahead	3M ahead	6M ahead	9M ahead	12M ahead
Buying Conditions	1.00	0.98	0.96	0.96*	0.97**	0.98**
ISM New Orders	0.99	1.00	0.99	0.99	0.98	0.97
Housing starts	0.99	0.99	1.00	0.97	0.95	0.94
ADS index	1.03	1.00	1.00	0.99	1.00	1.00
Term spread	1.03	1.03	0.99*	0.91***	0.88***	0.87***
Default spread	1.09	1.13	1.14	1.20	1.22	1.20
3M T-bill rate	1.01	1.02	1.02	1.02	1.02	1.01
Excess market return	1.08	1.03	1.02	1.04	1.06**	1.08**
Realised volatility	1.14*	1.23*	1.22	1.29	1.32	1.31
First principal component	0.99	1.00	1.00	1.02	1.04	1.02
Second principal component	1.01	0.99	0.98	0.96*	0.96	0.95*
Third principal component	1.05	1.00	0.99	0.98*	0.98	0.99

Benchmark: $GJR-GARCH(1,1)$. MAFE ratio: $\frac{MAFE_{GMX}}{MAFE_{GARCH}}$, where $MAFE_{GMX}$ stands for the mean absolute forecast error from the GARCH-MIDAS model driven by some macroeconomic or financial data (X). A value below 1 means the GARCH-MIDAS model outperforms the $GJR-GARCH(1,1)$ model. *, ** and *** indicate a rejection of the null hypothesis of equal (conditional) predictive ability at the 10%, 5%, and 1% level, respectively, according to the Giacomini and White (2006) test. Conditioning variable: industrial production growth (manufacturing only, final release). Test function: $h_t = [1 \ v_t]$, where v_t is the conditioning information. $RV_t = \sum_{i=1}^{N_t} |r_{i,t}|$.

Industrial production data and the NBER recession dates largely replicate the results in Table A.15. Conditioning on the market stress indicator leads to slightly stronger results than when conditioning on the VIX index (Table 7.

Table A.17: Conditional test using market stress indicator (level): MAFE ratios

	1M ahead	2M ahead	3M ahead	6M ahead	9M ahead	12M ahead
Buying Conditions	1.00	0.98	0.96	0.96**	0.97**	0.98**
ISM New Orders	0.99	1.00	0.99	0.99**	0.98**	0.97*
Housing starts	0.99	0.99	1.00	0.97**	0.95**	0.94*
ADS index	1.03	1.00	1.00	0.99	1.00	1.00
Term spread	1.03	1.03	0.99	0.91***	0.88***	0.87***
Default spread	1.09	1.13	1.14	1.20	1.22	1.20
3M T-bill rate	1.01	1.02*	1.02***	1.02***	1.02***	1.01**
Excess market return	1.08	1.03	1.02	1.04*	1.06**	1.08**
Realised volatility	1.14	1.23*	1.22	1.29	1.32	1.31
First principal component	0.99	1.00	1.00	1.02	1.04	1.02
Second principal component	1.01	0.99	0.98	0.96*	0.96**	0.95**
Third principal component	1.05	1.00	0.99	0.98*	0.98*	0.99*

Benchmark: $GJR-GARCH(1,1)$. MAFE ratio: $\frac{MAFE_{GMX}}{MAFE_{GARCH}}$, where $MAFE_{GMX}$ stands for the mean absolute forecast error from the GARCH-MIDAS model driven by some macroeconomic or financial data (X). A value below 1 means the GARCH-MIDAS model outperforms the $GJR-GARCH(1,1)$ model. *, ** and *** indicate a rejection of the null hypothesis of equal (conditional) predictive ability at the 10%, 5%, and 1% level, respectively, according to the Giacomini and White (2006) test. Conditioning variable: level of financial stress index. Test function: $h_t = [1 \ v_t]$, where v_t is the conditioning information. $RV_t = \sum_{i=1}^{N_t} |r_{i,t}|$.

Table A.18: Conditional test using past forecast loss differences: MAFE ratios

	1M ahead	2M ahead	3M ahead	6M ahead	9M ahead	12M ahead
Buying Conditions	1.00	0.98	0.96	0.96	0.97**	0.98**
ISM New Orders	0.99	1.00	0.99	0.99	0.98	0.97
Housing starts	0.99*	0.99	1.00	0.97	0.95	0.94
ADS index	1.03	1.00	1.00	0.99	1.00	1.00
Term spread	1.03	1.03	0.99	0.91***	0.88***	0.87***
Default spread	1.09	1.13	1.14	1.20	1.22	1.20
3M T-bill rate	1.01	1.02	1.02***	1.02**	1.02***	1.01*
Excess market return	1.08	1.03	1.02	1.04	1.06**	1.08*
Realised volatility	1.14	1.23*	1.22	1.29	1.32	1.31
First principal component	0.99	1.00	1.00	1.02	1.04	1.02
Second principal component	1.01	0.99	0.98	0.96	0.96	0.95
Third principal component	1.05*	1.00	0.99	0.98	0.98	0.99

Benchmark: $GJR-GARCH(1,1)$. MAFE ratio: $\frac{MAFE_{GMX}}{MAFE_{GARCH}}$, where $MAFE_{GMX}$ stands for the mean absolute forecast error from the GARCH-MIDAS model driven by some macroeconomic or financial data (X). A value below 1 means the GARCH-MIDAS model outperforms the $GJR-GARCH(1,1)$ model. *, ** and *** indicate a rejection of the null hypothesis of equal (conditional) predictive ability at the 10%, 5%, and 1% level, respectively, according to the Giacomini and White (2006) test. Conditioning variable: forecast loss difference at the forecast origin. Test function: $h_t = [1 \ v_t]$, where v_t is the conditioning information. $RV_t = \sum_{i=1}^{N_t} |r_{i,t}|$.

A.10 Combination forecasts of all the models

This appendix presents results for the combination forecasts, where also the financial data is included in the combinations. These combinations are thus based on a larger set of data than those in Section 7. As expected, because the forecasts produced by the GARCH-MIDAS models driven by financial data perform poorly in almost all periods and on almost all horizons, including them in the forecast combinations mainly leads to clearly inferior performance.

Table A.19: Combination forecasts: MAFE ratios

	1 month ahead	3 month ahead	6 month ahead	9 month ahead	12 month ahead
Mean	1.01	0.99	1.00	1.01	1.00
Median	0.98	0.98	0.97	0.97	0.96*
Trimmed mean	1.00	0.99	0.99	0.99	0.98
DMAPE, $\eta = 1$	1.00	0.98	1.00	1.01	0.99
DMAPE, $\eta = 0.5$	1.01	0.97	0.99	1.00	0.99
Previously best	1.11	1.01	0.95	0.96	0.89***
Mean (best three)	1.02	0.96	0.92***	0.93**	0.94*

Benchmark: GJR-GARCH(1,1). MAFE ratio: $\frac{MAFE_{combo}}{MAFE_{GARCH}}$, where $MAFE_{combo}$ stands for the mean absolute forecast error from the combination forecast using the method stated in the first column. A value below 1 means the combination forecast outperforms the GJR-GARCH(1,1) model. *, ** and *** indicate a rejection of the null hypothesis of equal (unconditional) predictive ability at the 10%, 5%, and 1% level, respectively, according to the Giacomini and White (2006) test. The last four combination schemes are based on the forecasting performance over an expanding window of initial size 12 months. Note that due to initial calculations all forecast comparisons are for the period January 1998 - June 2017 (234 periods).

Table A.20: Combination forecasts: MSFE ratios

	1 month ahead	3 month ahead	6 month ahead	9 month ahead	12 month ahead
Mean	1.03	0.97	0.97	0.98	0.99
Median	0.99	0.97	0.96	0.96	0.98*
Trimmed mean	1.02	0.97	0.97	0.97	0.98
DMSPE, $\eta = 1$	1.03	0.96	0.97	0.98	0.98
DMSPE, $\eta = 0.5$	1.05	0.96	0.97	0.98	0.98
Previously best	1.40	0.98	0.95	0.92	0.93
Mean (best three)	1.05	0.97	0.92	0.92	0.95*

Benchmark: GJR-GARCH(1,1). MSFE ratio: $\frac{MSFE_{combo}}{MSFE_{GARCH}}$, where $MSFE_{combo}$ stands for the mean squared forecast error from the combination forecast using the method stated in the first column. A value below 1 means the combination forecast outperforms the GJR-GARCH(1,1) model. *, ** and *** indicate a rejection of the null hypothesis of equal (unconditional) predictive ability at the 10%, 5%, and 1% level, respectively, according to the Giacomini and White (2006) test. Note that due to initial training periods etc., all forecast comparisons are for the period January 1998 - June 2017 (234 periods).

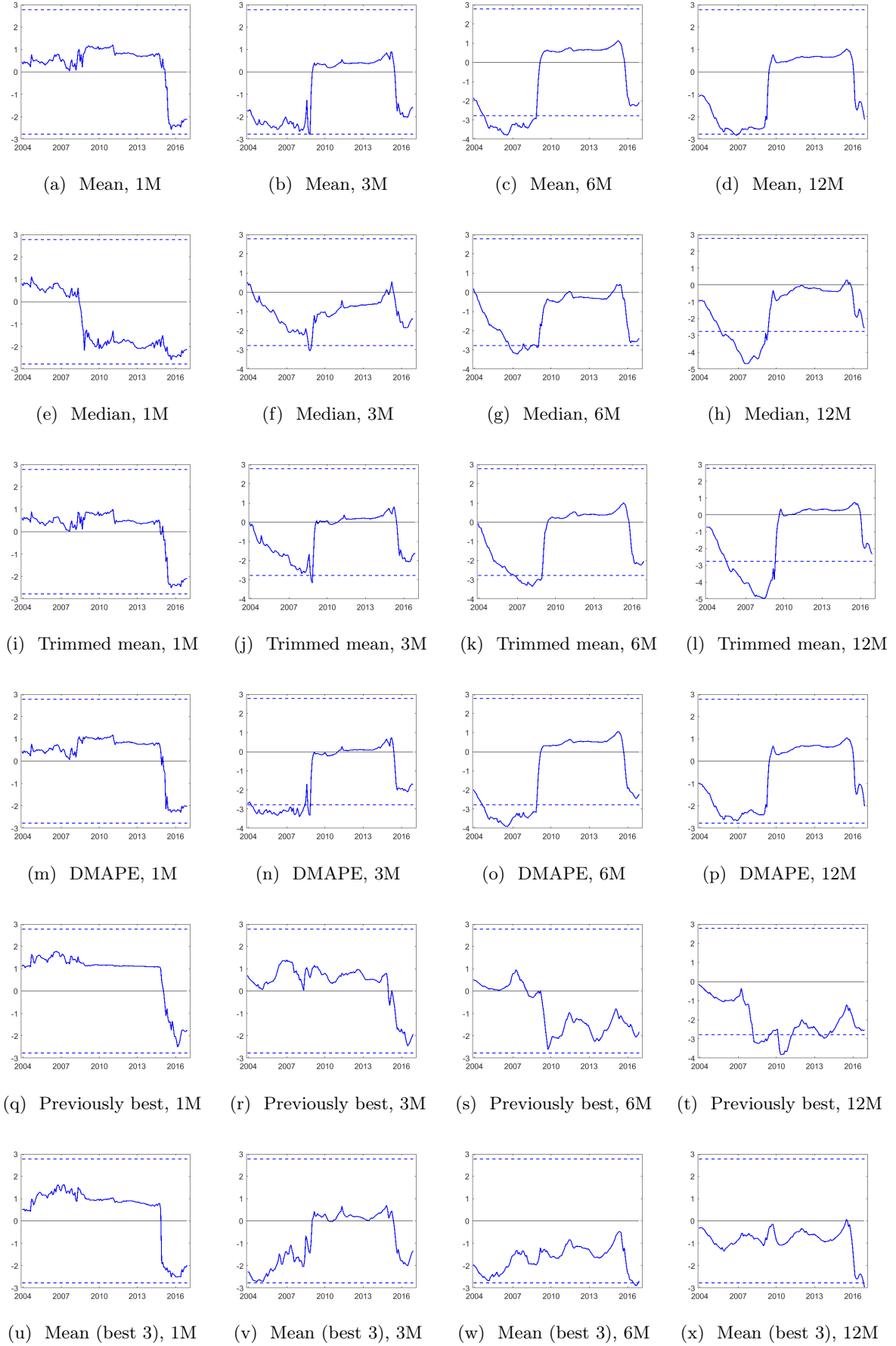


Figure A.24: Fluctuation test applied to forecast combinations of the individual GARCH-MIDAS models. Dashed lines represent 10% confidence bands. Benchmark: GJR-GARCH(1,1) model. DMAPE with $\eta = 0.5$. Note that the year on the x-axis marks the end of the rolling window period, over which the test statistics is calculated. $l = 5$, $m = 78$

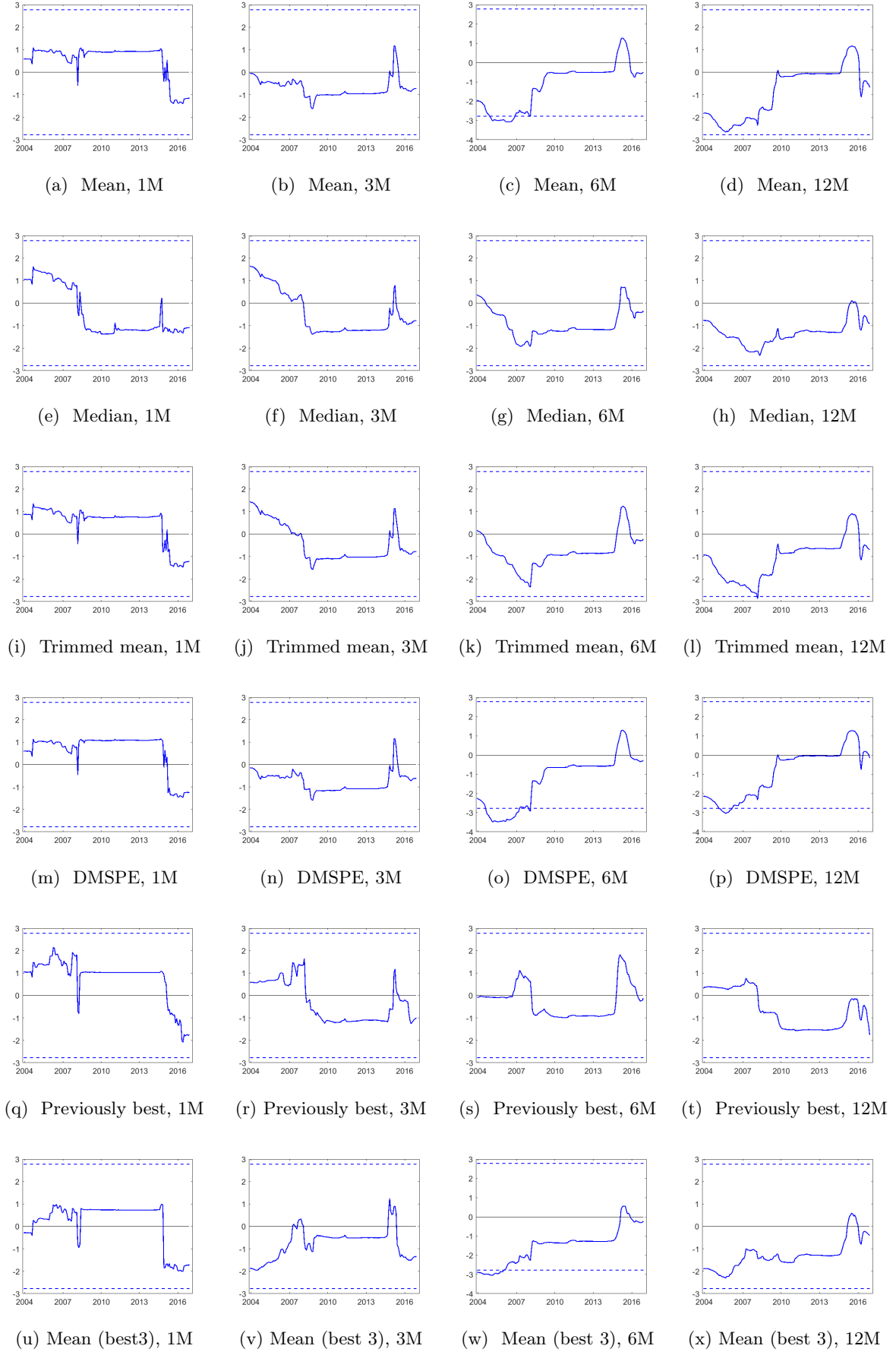


Figure A.25: Fluctuation test applied to forecast combinations of the individual GARCH-MIDAS models. Squared forecast errors. Dashed lines represent 10% confidence bands. Benchmark: GJR-GARCH(1,1) model. DMSPE with $\eta = 0.5$. Note that the year on the x-axis marks the end of the rolling window period, over which the test statistics is calculated. $l = 5$, $m = 78$

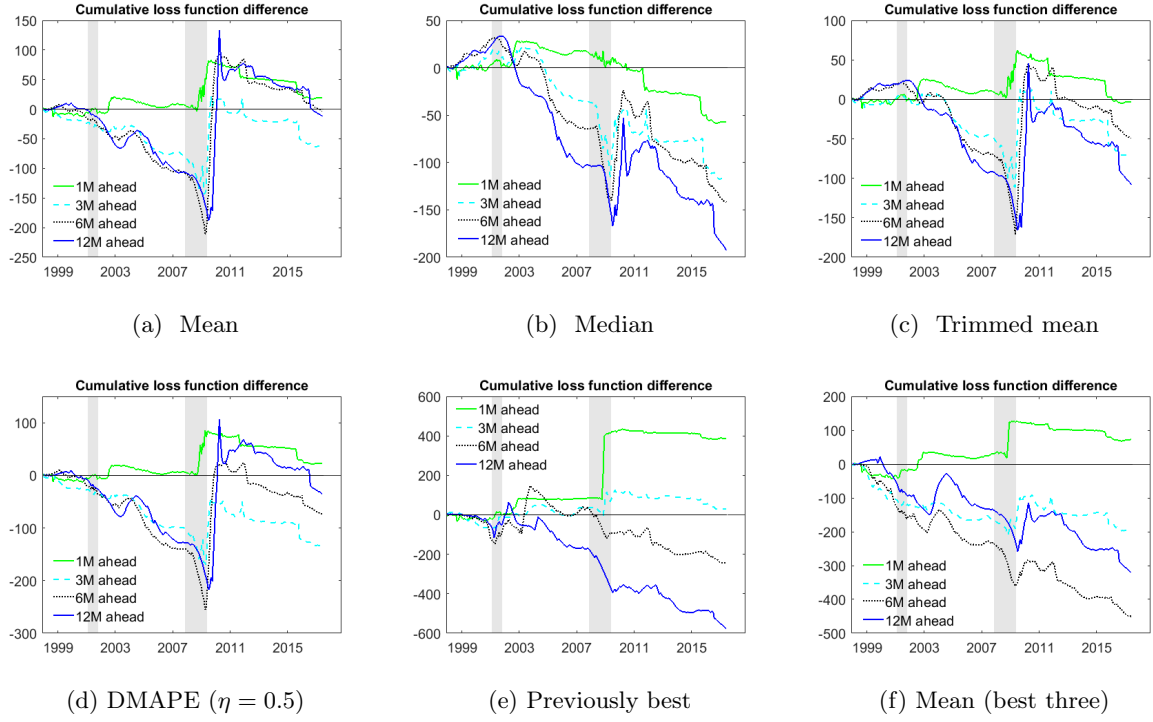


Figure A.26: Cumulative sum of loss function differences (absolute errors) of forecast combinations of the individual GARCH-MIDAS models, compared to the GJR-GARCH(1,1) model ($|Loss_{combo} - Loss_{GARCH}|$). An upward sloping segment thus indicates the GJR-GARCH model outperforms the GARCH-MIDAS model. Grey areas mark NBER dated US recessions.

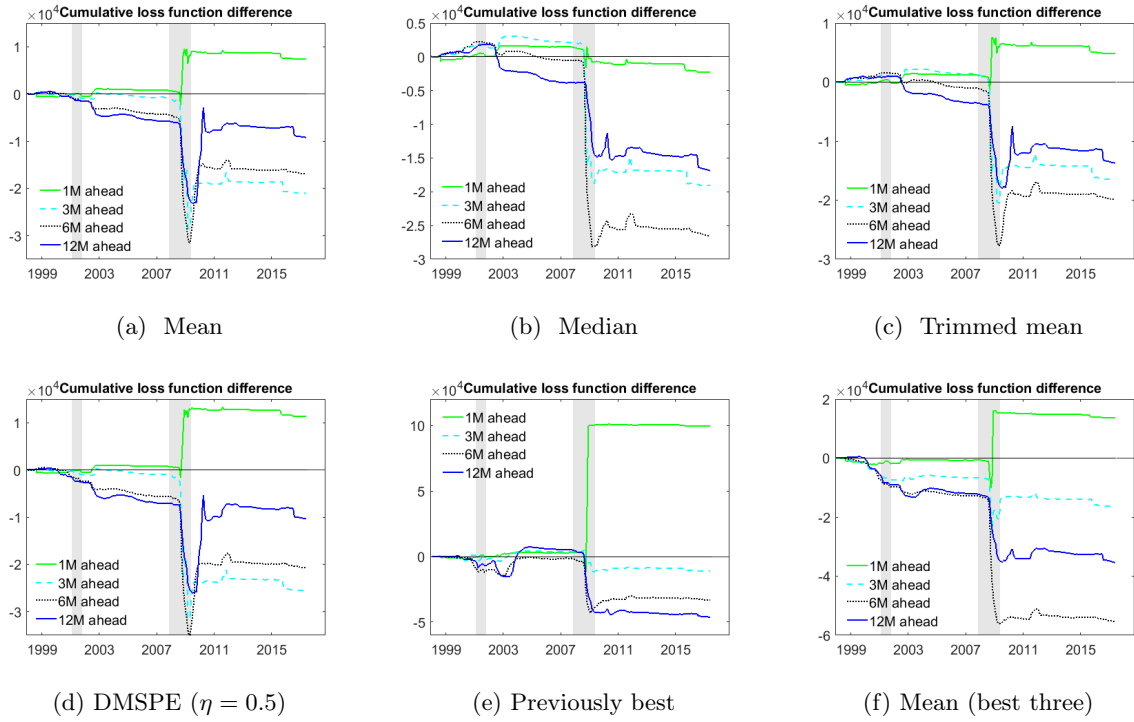


Figure A.27: Cumulative sum of loss function differences (squared errors) of forecast combinations of the individual GARCH-MIDAS models, compared to the GJR-GARCH(1,1) model ($(Loss_{combo} - Loss_{GARCH})^2$). An upward sloping segment thus indicates the GJR-GARCH model outperforms the GARCH-MIDAS model. Grey areas mark NBER dated US recessions.

A.11 Mincer-Zarnowitz regressions

To check whether the forecasts are unbiased I present results for the Mincer-Zarnowitz regressions for selected horizons. They take the following form: $RV_t = c + \rho \widehat{RV}_{t,t-h} + \epsilon_t$, where $\widehat{RV}_{t,t-h}$ is the forecasts from the GARCH model or GARCH-MIDAS model. If the forecast is unbiased $c = 0$ and $\rho = 1$. In Table ?? I test both the individual hypothesis that $c = 0$ or $\rho = 1$, as well as use an F-test to test whether c and ρ are jointly equal to 0 and 1, respectively.

Clearly, for the 1 month and 3 months ahead forecasts we cannot reject the null hypothesis of $c = 0$ and $\rho = 1$, and thus the forecasts seem unbiased. For the 6 months ahead horizon the forecasts produced by the macroeconomic data, the term spread and the principal components seem unbiased. On the 12 months horizon only the GARCH-MIDAS models driven by housing starts, the term spread and the second principal component seem to produce unbiased forecasts. Looking at the (unadjusted) R^2 statistics performance between the models diverges quite significantly at the longer horizons, with especially the term spread and housing starts driven GARCH-MIDAS models performing well.

Table A.21: Mincer-Zarnowitz regression estimates

	1M ahead				3M ahead				6M ahead				12M ahead			
	c	ρ	F stat	R^2	c	ρ	F stat	R^2	c	ρ	F stat	R^2	c	ρ	F stat	R^2
GJR-GARCH(1,1)	-2.85 (5.30)	1.30 (0.25)	456***	64.04	6.07 (4.42)	1.17 (0.24)	25.2***	8.95	15.04 (12.13)	0.82 (0.75)	2.04	0.79	53.73*** (12.70)	-1.28*** (0.56)	2.19	0.85
Realised volatility	4.97 (4.10)	0.84 (0.20)	273***	51.59	20.99*** (4.90)	0.33*** (0.09)	17.8***	6.49	27.32*** (6.75)	0.11*** (0.06)	1.93	0.75	30.26*** (7.21)	0.01*** (0.05)	0.01	0.00
Buying Conditions index	-3.83 (5.21)	1.36 (0.25)	509***	66.52	-4.97 (8.04)	1.53* (0.50)	68.1***	21.01	-1.43 (13.15)	1.63 (0.82)	29.4***	10.29	22.00*** (7.90)	0.44 (0.40)	1.76	0.68
ISM New Orders index	-2.45 (5.50)	1.32 (0.27)	477***	65.06	3.85 (4.86)	1.35 (0.33)	29.8***	10.42	12.51 (10.91)	0.99 (0.76)	4.76**	1.83	24.70*** (8.96)	0.33* (0.39)	0.42	0.16
Housing starts	-3.83 (4.88)	1.35 (0.24)	496***	65.94	1.56 (5.43)	1.37 (0.36)	42.9***	14.34	-2.69 (16.03)	1.68 (0.98)	27.3***	9.64	-5.21 (22.13)	1.86 (1.36)	28.5***	10.03
ADS index	-2.85 (5.76)	1.33 (0.29)	378***	59.59	-0.11 (6.97)	1.47 (0.45)	44.0***	14.66	5.43 (10.78)	1.30 (0.72)	12.1***	4.52	22.62*** (7.37)	0.42 (0.44)	0.97	0.38
Term spread	-10.42* (5.60)	1.79*** (0.29)	499***	66.11	-22.83 (15.29)	2.89* (0.98)	57.9***	18.45	-33.50 (26.17)	3.64 (1.67)	56.9***	18.19	-31.51 (23.64)	3.58 (1.57)	56.6***	18.10
Default spread	7.68 (4.91)	0.82 (0.23)	294***	53.47	20.10*** (4.42)	0.44*** (0.18)	19.0***	6.92	27.68*** (6.16)	0.12*** (0.10)	1.30	0.51	31.10*** (6.90)	-0.03*** (0.05)	0.06	0.02
3M T-bill rate	-2.13 (5.14)	1.38 (0.26)	468***	64.62	8.48* (4.37)	1.22 (0.28)	23.2***	8.32	21.20* (12.57)	0.59 (0.99)	0.93	0.36	36.30*** (12.66)	-0.39 (1.07)	0.24	0.09
Excess market return	-3.74 (7.64)	1.33 (0.39)	248***	49.25	9.71* (5.72)	0.93 (0.30)	18.1***	6.61	31.78** (15.70)	-0.06* (0.61)	0.02	0.01	48.61*** (15.15)	-0.96*** (0.56)	5.42**	2.07
First principal component	-1.36 (4.78)	1.22 (0.22)	528***	67.33	7.63* (4.57)	1.04 (0.27)	33.2***	11.48	16.48** (8.22)	0.68 (0.49)	6.80***	2.59	24.42*** (7.03)	0.30** (0.31)	1.14	0.44
Second principal component	-2.70 (4.84)	1.42* (0.25)	493***	65.81	0.85 (5.11)	1.66* (0.37)	42.3***	14.17	-14.52 (23.23)	2.83 (1.66)	30.2***	10.57	0.65 (15.40)	1.96 (1.25)	9.17***	3.46
Third principal component	5.05 (6.01)	1.05 (0.31)	285***	52.70	11.71*** (3.76)	0.98 (0.20)	29.5***	10.35	14.52* (7.65)	0.92 (0.55)	7.17***	2.72	24.19*** (8.78)	0.38 (0.61)	0.74	0.29

Parameter estimates for MZ regressions, estimated using OLS. HAC standard errors can be found in the parenthesis. *, ** and *** indicate a rejection of the null hypothesis (i.e., that $c = 0$ or $\rho = 1$, or for the F statistic that jointly $c = 0$ and $\rho = 1$) at the 10%, 5%, and 1% level, respectively.